PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

Rec'd PCT/PTO 20 DEC 2005

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7: C12N 15/29

A1

(11) International Publication Number: WO 00/42190

(43) International Publication Date: 20 July 2000 (20.07.00)

(21) International Application Number: PCT/AU00/00007

(22) International Filing Date: 7 January 2000 (07.01.00)

(30) Priority Data:
PP-8077 8 January 1999 (08.01.99) AU

(71) Applicant *(for all designated States except US)*: THE UNIVERSITY OF QUEENSLAND [AU/AU]; St Lucia, QLD 4067 (AU).

(71) Applicant *(for US only)*: SUN, Xiao-Yi (executrix for the deceased inventor) [AU/AU]; 21 Cooranga Street, Jindalee, QLD 4074 (AU).

(72) Inventor: ZHOU, Jian (deceased).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: FRAZER, Ian, Hector [AU/AU]; 64 Eighth Avenue, St Lucia, QLD 4072 (AU). BOTELLA MESA, Jose, Ramon [ES/AU]; 12 Tad Street, Kenmore, QLD 4069 (AU).

(74) Agents: ARGAET, Victor, P. et al.; Davies Collison Cave, 1 Little Collins Street, Melbourne, VIC 3000 (AU).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: POLYNUCLEOTIDE AND METHOD FOR SELECTIVELY EXPRESSING A PROTEIN IN A TARGET CELL OR TISSUE OF A PLANT

(57) Abstract

A method is disclosed for constructing a synthetic polynucleotide from which a protein is selectively expressible in a target cell of a plant, relative to another cell of the plant. The method comprises selecting a first codon of a parent polynucleotide for replacement with a synonymous codon which has a higher translational efficiency in the target cell than in said other cell, and replacing said first codon with said synonymous codon to form said synthetic polynucleotide.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

POLYNUCLEOTIDE AND METHOD FOR SELECTIVELY EXPRESSING A PROTEIN IN A TARGET CELL OR TISSUE OF A PLANT

FIELD OF THE INVENTION

THIS INVENTION relates generally to gene expression and in particular, to modulation of gene expression by changing codon composition of a polynucleotide. More particularly, the present invention relates to synthetic polynucleotides and methods for selectively expressing a protein in a target cell or tissue of a plant in which at least one existing codon of a parent polynucleotide encoding the protein has been replaced with a synonymous codon.

BACKGROUND OF THE INVENTION

Selective targeting of genes to particular plant cells or tissues would have broad utility for producing transgenic plants with desirable novel phenotypes. However, due to a lack of tissue-specific promoters currently available, selective expression of genes in target cells or tissues of a plant has proven extremely difficult. Accordingly, there is a need to provide technologies which can facilitate selective expression of proteins to a particular target cell or tissue of a plant.

SUMMARY OF THE INVENTION

In one aspect of the invention, there is provided a method of constructing a synthetic polynucleotide from which a protein is selectively expressible in a target cell of a plant, relative to another cell of the plant, said method comprising:

- selecting a first codon of a parent polynucleotide for replacement with a synonymous codon which has a higher

translational efficiency in said target cell than in said other cell; and

- replacing said first codon with said synonymous codon to form said synthetic polynucleotide.

Preferably, said first codon and said synonymous codon are selected by:

- comparing translational efficiencies of individual codons in said target cell relative to said other cell; and

- selecting said first codon and said synonymous codon based on said measurement.

A translational efficiency of a codon may be determined by any suitable technique. In a preferred embodiment, the translational efficiency of a codon is measured by:

- introducing into said target cell and into said other cell, a synthetic construct comprising a reporter polynucleotide fused in frame with a tandem repeat of said individual codon, wherein said reporter polynucleotide encodes a reporter protein, and wherein said synthetic construct is operably linked to a regulatory polynucleotide; and

- comparing expression of said reporter protein in said target cell and in said other cell to determine the translational efficiency of said individual codon in said target cell relative to said other cell.

Preferably, the above method is further characterized by:

- introducing the synthetic construct into a progenitor cell of a cell selected from the group consisting of said target cell and said another cell; and

- producing said cell from said progenitor cell, wherein said cell contains said synthetic construct.

Suitably, this method is further characterized by:

- introducing the synthetic construct into a progenitor cell of a cell selected from the group consisting of said target cell and said another cell; and

- growing a plant or part thereof from said progenitor cell, wherein said plant comprises said cell containing said synthetic construct.

The above method may be further characterized by the step of introducing the synthetic construct into a plant or part thereof such that said synthetic construct is introduced into said target cell or said other cell.

Preferably, said synonymous codon corresponds to a reporter construct from which the reporter protein is expressed in said target cell at a level that is at least 110%, preferably at least 200%, more preferably at least 500%, and most preferably at least 1000%, of that expressed from the said reporter construct in said other cell.

In an alternate embodiment, the translational efficiency of a codon may be determined by measuring the abundance of an iso-tRNA corresponding to said individual codon in said target cell relative to said other cell.

Preferably, said synonymous codon corresponds to an iso-tRNA which is in higher abundance in the target cell relative to said other cell.

Preferably, selecting said first codon and said synonymous codon comprises:

- measuring abundance of different iso-tRNAs in said target cell relative to said other cell; and

- selecting said first codon and said synonymous codon based on said measurement, wherein said synonymous codon corresponds to an iso-tRNA which is in higher abundance in said target cell than in said other cell.

Advantageously, said synonymous codon corresponds to an iso-tRNA that is present in said target cell at a level which is at least 110%, preferably at least 200%, more preferably at least 500%, and most preferably at least 1000%, of the level that is present in said other cell.

Alternatively, the step of selecting may be characterized in that a synonymous codon according to the invention is selected from the group consisting of (1) a codon used at relatively high frequency by genes, preferably highly expressed genes, of a said target cell or tissue, (2) a codon used at relatively high frequency by genes, preferably highly expressed genes, of the plant, (3) a codon used at relatively low frequency by genes of a said one or more other cells or tissues, and (4) a codon used at relatively low frequency by genes of another organism.

The step of selecting may be characterized in that a first codon according to the invention is selected from the group consisting of (a) a codon used at relatively high frequency by genes, preferably highly expressed genes, of a said one or more other cells or tissues, (b) a codon used at relatively low frequency by genes of a said target cell or tissue, (c) a codon used at relatively low frequency by genes of the plant, and (d) a codon used at relatively high frequency by genes of another organism.

In a preferred embodiment, the method further includes the step of selecting the first codon and the synonymous codon such that said protein is expressed from said synthetic polynucleotide in said target cell or tissue at a

level which is at least 110%, preferably at least 200%, more preferably at least 500%, and most preferably at least 1000%, of that expressed from said parent polynucleotide in said target cell or tissue.

Preferably, the other cell is a precursor cell of the target cell. Alternatively, the other cell may be a cell derived from the target cell.

In another aspect, the invention provides a synthetic polynucleotide constructed according to any one of the above methods.

In yet another aspect, the invention resides in a method for selectively expressing a protein in a target cell or tissue of a plant, said method comprising:

- replacing a first codon of a parent polynucleotide encoding said protein with a synonymous codon to produce a synthetic polynucleotide having altered translational kinetics compared to said parent polynucleotide, such that said protein is expressible in said target cell, but such that said protein is not substantially expressible in another cell of the plant; and

- introducing into a cell selected from the group consisting of said target cell and a precursor of said target cell, said synthetic polynucleotide operably linked to a regulatory polynucleotide. The protein is thereby selectively expressed in said target cell.

Preferably, said synonymous codon has a higher translational efficiency in said target cell than in said other cell.

In yet another aspect, the invention provides a method of expressing a protein in a target cell of a plant from a first polynucleotide, said method comprising:

- introducing into a cell selected from the group consisting of said target cell and a precursor of said target cell, a second polynucleotide encoding an iso-tRNA, wherein said second polynucleotide is operably linked to a regulatory polynucleotide, and wherein said iso-tRNA is normally in relatively low abundance in said target cell and corresponds to a codon of said first polynucleotide. The protein is thereby expressed in the target cell.

The invention further contemplates cells or tissues containing therein the synthetic polynucleotides of the invention, or alternatively, cells or tissues produced from the methods of the invention.

In yet another aspect, the invention resides in a transgenic plant or plant part containing cells or tissues having the synthetic polynucleotides of the invention.

DETAILED DESCRIPTION

The present invention is based, at least in part, on the discovery in copending International Application No. PCT/AU98/00530 that the intracellular abundance of different species of isoaccepting transfer RNAs (iso-tRNAs) in different cell or tissue types contributes to the expression of messenger RNA (mRNA) by determining the rate of protein translation from the mRNA. By altering the structure of a protein-encoding polynucleotide to replace existing codons with synonymous codons that correspond to iso-tRNAs that are in higher or lower abundance intracellularly, it was shown that the translational efficiency of a mRNA can be dramatically modulated both *in vitro* and *in vivo*. This ability to control the translational kinetics of a modified mRNA allows for the modified mRNA to be selectively expressible in a particular cell or tissue, or alternatively

in a cell or tissue in a specific state of differentiation or in a specific stage of the cell cycle.

Although the subject matter of PCT/AU98/00530 is primarily concerned with selective expression of a protein in a target cell or tissue of a mammal, the present inventors believe that different species of iso-tRNAs would be differentially expressed in different cells or tissues of most multi-cellular organism. Put another way, the inventors consider that the translational efficiencies of different codons will vary in different cells or tissues. Such differences can be exploited together with codon composition of a gene to regulate and direct expression of a protein to a particular cell or cell type of a plant, including cells in a selected tissue. Alternatively, these differences can be exploited together with codon composition of a gene to regulate and direct expression of a protein to a cell or tissue of a plant in a selected state of differentiation or in a selected stage of the cell cycle.

1. Definitions

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by those of ordinary skill in the art to which the invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, preferred methods and materials are described. For the purposes of the present invention, the following terms are defined below.

The articles "*a*" and "*an*" are used herein to refer to one or to more than one (i.e. to at least one) of the grammatical object of the article. By way of example, "an element" means one element or more than one element.

Throughout this specification, unless the context requires otherwise, the words *"comprise"*, *"comprises"* and *"comprising"* will be understood to imply the inclusion of a stated step or element or group of steps or elements but not the exclusion of any other step or element or group of steps or elements.

By *"expressible"* is meant expression of a protein to a level sufficient to effect a particular function associated with the protein. By contrast, the terms *"not expressible"* and *"not substantially expressible"* as used interchangeably herein refers to (a) no expression of a protein, (b) expression of a protein to a level that is not sufficient to effect a particular function associated with the protein, (c) expression of a protein, which cannot be detected by a monoclonal antibody specific for the protein, or (d) expression of a protein, which is less that 1% of the level expressed in a wild-type cell that normally expresses the protein.

By *"expressing said synthetic construct"* is meant transcribing the synthetic construct such that mRNA is produced.

By *"expression vector"* is meant any autonomous genetic element capable of directing the synthesis of a protein encoded by the vector. Such expression vectors are known by practitioners in the art.

As used herein, the term "function" refers to a biological, enzymatic, physical, chemical or therapeutic function.

The terms *"growing"* or *"regeneration"* as used herein mean growing a whole, differentiated plant from a plant cell, a group of plant cells, a plant part (including

seeds), or a plant piece (e.g., from a protoplast, callus, or tissue part).

By *"highly expressed genes"* is meant genes that express high levels of mRNA, and preferably high level of protein, relative to other genes.

By *"isoaccepting transfer RNA"* or *"iso-tRNA"* is meant one or more transfer RNA molecules that differ in their anticodon nucleotide sequence but are specific for the same amino acid.

By *"marker gene"* is meant a gene that imparts a distinct phenotype to cells expressing the marker gene and thus allow such transformed cells to be distinguished from cells that do not have the marker. A selectable marker gene confers a trait for which one can 'select' based on resistance to a selective agent (*e.g.*, a herbicide, antibiotic, radiation, heat, or other treatment damaging to untransformed cells). A screenable marker gene (or reporter gene) confers a trait that one can identify through observation or testing, *i.e.*, by 'screening' (*e.g.* β-glucuronidase, luciferase, or other enzyme activity not present in untransformed cells).

By *"natural gene"* is meant a gene that naturally encodes the protein. However, it is possible that the parent polynucleotide encodes a protein that is not naturally-occurring but has been engineered using recombinant techniques.

The term *"non-cycling cell"* as used herein refers to a cell that has withdrawn from the cell cycle and has entered the G0 state. In this state, it is known that transcription of endogenous genes and protein translation are at substantially reduced levels compared to phases of the cell cycle, namely G1, S, G2 and M. By contrast, the term

"*cycling cell*" as used herein refers to a cell, which is in one of the above phases of the cell cycle.

By "*obtained from*" is meant that a sample such as, for example, a polynucleotide extract or polypeptide extract is isolated from, or derived from, a particular source of the host. For example, the extract can be obtained from a tissue or a biological fluid isolated directly from the host.

The term "*oligonucleotide*" as used herein refers to a polymer composed of a multiplicity of nucleotide residues (deoxyribonucleotides or ribonucleotides, or related structural variants or synthetic analogues thereof) linked via phosphodiester bonds (or related structural variants or synthetic analogues thereof). Thus, while the term "oligonucleotide" typically refers to a nucleotide polymer in which the nucleotide residues and linkages between them are naturally occurring, it will be understood that the term also includes within its scope various analogues including, but not restricted to, peptide nucleic acids (PNAs), phosphoramidates, phosphorothioates, methyl phosphonates, 2-O-methyl ribonucleic acids, and the like. The exact size of the molecule can vary depending on the particular application. An oligonucleotide is typically rather short in length, generally from about 10 to 30 nucleotide residues, but the term can refer to molecules of any length, although the term "polynucleotide" or "nucleic acid" is typically used for large oligonucleotides.

By "*operably linked*" is meant that transcriptional and translational regulatory polynucleotides are positioned relative to a polypeptide-encoding polynucleotide in such a manner that the polynucleotide is transcribed and the polypeptide is translated.

As used herein, "*plant*" and "*differentiated plant*" refer to a whole plant or plant part containing differentiated plant cell types, tissues and/or organ systems. Plantlets and seeds are also included within the meaning of the foregoing terms. Plants included in the invention are any plants amenable to transformation techniques, including angiosperms, gymnosperms, monocotyledons and dicotyledons.

The term "*plant cell*" as used herein refers to protoplasts, gamete-producing cells, and cells which regenerate into whole plants. Plant cells include cells in plants as well as protoplasts in culture.

By "*plant tissue*" is meant differentiated and undifferentiated tissue derived from roots, shoots, pollen, seeds, tumor tissue, such as crown galls, and various forms of aggregations of plant cells in culture, such as embryos and calluses.

"*Polypeptide*", "*peptide*" and "*protein*" are used interchangeably herein to refer to a polymer of amino acid residues and to variants and synthetic analogues of the same. Thus, these terms apply to amino acid polymers in which one or more amino acid residues is a synthetic non-naturally occurring amino acid, such as a chemical analogue of a corresponding naturally occurring amino acid, as well as to naturally-occurring amino acid polymers.

The term "*polynucleotide*" or "*nucleic acid*" as used herein designates mRNA, RNA, cRNA, cDNA or DNA. The term typically refers to oligonucleotides greater than 30 nucleotide residues in length.

By "*primer*" is meant an oligonucleotide which, when paired with a strand of DNA, is capable of initiating the synthesis of a primer extension product in the presence of a

suitable polymerizing agent. The primer is preferably single-stranded for maximum efficiency in amplification but can alternatively be double-stranded. A primer must be sufficiently long to prime the synthesis of extension products in the presence of the polymerization agent. The length of the primer depends on many factors, including application, temperature to be employed, template reaction conditions, other reagents, and source of primers. For example, depending on the complexity of the target sequence, the oligonucleotide primer typically contains 15 to 35 or more nucleotide residues, although it can contain fewer nucleotide residues. Primers can be large polynucleotides, such as from about 200 nucleotide residues to several kilobases or more. Primers can be selected to be "substantially complementary" to the sequence on the template to which it is designed to hybridize and serve as a site for the initiation of synthesis. By "substantially complementary", it is meant that the primer is sufficiently complementary to hybridize with a target polynucleotide. Preferably, the primer contains no mismatches with the template to which it is designed to hybridize but this is not essential. For example, non-complementary nucleotide residues can be attached to the 5′ end of the primer, with the remainder of the primer sequence being complementary to the template. Alternatively, non-complementary nucleotide residues or a stretch of non-complementary nucleotide residues can be interspersed into a primer, provided that the primer sequence has sufficient complementarity with the sequence of the template to hybridize therewith and thereby form a template for synthesis of the extension product of the primer.

"*Probe*" refers to a molecule that binds to a specific sequence or sub-sequence or other moiety of another

molecule. Unless otherwise indicated, the term "probe" typically refers to a polynucleotide probe that binds to another polynucleotide, often called the "target polynucleotide", through complementary base pairing. Probes can bind target polynucleotides lacking complete sequence complementarity with the probe, depending on the stringency of the hybridization conditions. Probes can be labeled directly or indirectly.

The terms *"precursor cell or tissue"* and *"progenitor cell or tissue"* as used herein refer to a cell or tissue that can gives rise to a particular cell or tissue in which protein expression is to be targeted or in which translational efficiency of a codon is to be determined.

By *"recombinant polypeptide"* is meant a polypeptide made using recombinant techniques, *i.e.*, through the expression of a recombinant or synthetic polynucleotide.

"Stringency" as used herein, refers to the temperature and ionic strength conditions, and presence or absence of certain organic solvents, during hybridization. The higher the stringency, the higher will be the degree of complementarity between immobilized polynucleotides and the labeled polynucleotide.

"Stringent conditions" refers to temperature and ionic conditions under which only polynucleotides having a high frequency of complementary bases will hybridize. The stringency required is nucleotide sequence dependent and depends upon the various components present during hybridization. Generally, stringent conditions are selected to be about 10 to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic

strength and pH) at which 50% of a target sequence hybridizes to a complementary probe.

The term *"synthetic polynucleotide"* as used herein refers to a polynucleotide formed *in vitro* by the manipulation of a polynucleotide into a form not normally found in nature. For example, the synthetic polynucleotide can be in the form of an expression vector. Generally, such expression vectors include transcriptional and translational regulatory polynucleotide operably linked to the polynucleotide.

The term *"synonymous codon"* as used herein refers to a codon having a different nucleotide sequence than another codon but encoding the same amino acid as that other codon.

By *"translational efficiency"* is meant the efficiency of a cell's protein synthesis machinery to incorporate the amino acid encoded by a codon into a nascent polypeptide chain. This efficiency can be evidenced, for example, by the rate at which the cell is able to synthesize the polypeptide from an RNA template comprising the codon, or by the amount of the polypeptide synthesized from such a template.

The term *"transformation"* means alteration of the genotype of a host plant by the introduction of a chimeric nucleic acid.

By *"transgenote"* is meant an immediate product of a transformation process.

By *"vector"* is meant a polynucleotide molecule, preferably a DNA molecule derived, for example, from a plasmid, bacteriophage, or plant virus, into which a polynucleotide can be inserted or cloned. A vector preferably contains one or more unique restriction sites and can be capable of autonomous replication in a defined host

cell including a target cell or tissue or a progenitor cell or tissue thereof, or be integrable with the genome of the defined host such that the cloned sequence is reproducible. Accordingly, the vector can be an autonomously replicating vector, *i.e.*, a vector that exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, *e.g.*, a linear or closed circular plasmid, an extrachromosomal element, a minichromosome, or an artificial chromosome. The vector can contain any means for assuring self-replication. Alternatively, the vector can be one which, when introduced into the host cell, is integrated into the genome and replicated together with the chromosome(s) into which it has been integrated. A vector system can comprise a single vector or plasmid, two or more vectors or plasmids, which together contain the total DNA to be introduced into the genome of the host cell, or a transposon. The choice of the vector will typically depend on the compatibility of the vector with the host cell into which the vector is to be introduced. The vector can also include a selection marker such as an antibiotic resistance gene that can be used for selection of suitable transformants. Examples of such resistance genes are known to those of skill in the art and include the *nptII* gene that confers resistance to the antibiotics kanamycin and G418 (Geneticin®) and the *hph* gene which confers resistance to the antibiotic hygromycin B.

2. Selection of synonymous codons

According to the present invention, selective targeting of protein expression to a plant cell is effected by replacing at least one existing codon (sometimes referred to as a "first codon") of a parent polynucleotide encoding the

protein with a synonymous codon (i.e. one which encodes the same amino acid residue as the first codon).

Replacement of synonymous codons for existing codons is not new *per se*. In this regard, we refer to International Application Publication No WO 96/09378 which utilizes such substitution to provide a method of expressing proteins of eukaryotic and viral origin at high levels in *in vitro* mammalian cell culture systems, the main thrust of that method being the harvesting of such proteins. In distinct contrast, the present invention utilizes substitution of one or more codons in a gene for targeting expression of the gene to particular cells or tissues with the ultimate aim of producing transgenic plants with novel phenotypes.

The present method preferably includes the step of selecting the codons such that the synonymous codon has a higher translational efficiency in said target cell or tissue ("cell or tissue" is sometimes referred to herein as "cell/tissue") relative to said one or more other cells or tissues. As used herein, expression of a protein in a tissue refers alternatively to expression of the protein within a cell of the tissue or production of the protein within a cell and export of the protein from the cell to, for example, the extracellular matrix of a tissue.

Methods for determining translational efficiencies of different codons in and between different cells or tissues are described in detail in Section 3. The translational efficiencies so determined can be used to identify which isocoding triplets are differentially translated between the different cells or tissues. In a typical scenario, there will be: (A) codons with higher translational efficiencies in a target cell/tissue relative to one or more other cells/tissues; (B) codons with higher translational efficiencies in the one or more other cells/tissues relative

to the target cell/tissue; and (C) codons with about the same translational efficiencies in the target cell/tissue relative to the one or more other cells/tissues. Synonymous codons are selected such that they correspond to (A) codons. Preferably, a synonymous codon is selected such that it has the largest difference in translational efficiency in the target cell or tissue relative to the existing codon that it replaces. Existing codons in a parent polynucleotide are preferably selected such that they do not have the same translational bias as the synonymous codons with respect to the target cell/tissue and the one or more other cell/tissue (*i.e.*, existing codons should preferably not correspond to (A) codons). However, existing codons can have similar translational efficiencies in each of the target cell/tissue and the one or more other cells/tissues (*i.e.*, existing codons can correspond to (C) codons. They can also have a translational bias opposite to that of the synonymous codons (*i.e.*, existing codons can, and preferably do, correspond to (B) codons).

Suitably, a synonymous codon has a translational efficiency in the target cell/tissue that is at least 110%, preferably at least 200%, more preferably at least 500%, and still more preferably at least 1000%, of that in the other cell(s)/tissue(s). In the case of two or more synonymous codons having similar translational efficiencies in the target cell/tissue relative to the other cell(s)/tissue(s), it will be appreciated that any one of these codons can be used to replace the existing codon.

It is preferable but not necessary to replace all the existing codons of the parent polynucleotide with synonymous codons having higher translational efficiencies in the target cell/tissue compared to the other cells/tissues. Increased expression can be accomplished even with partial

replacement. Suitably, the replacement step affects 5%, 10%, 15%, 20%, 25%, 30%, more preferably 35%, 40%, 50%, 60%, 70% or more of the existing codons of the parent polynucleotide.

The difference in level of protein expressed in the target cell/tissue from a synthetic polynucleotide relative to that expressed in the other cell(s)/tissue(s) depends on the percentage of existing codons replaced by synonymous codons, and the difference in translational efficiencies of the synonymous codons in the target cell/tissue relative to the other cell(s)/tissue(s). Put another way, the fewer such replacements, and/or the smaller the difference in translational efficiencies of the synonymous between the different cells/tissues, the smaller the difference in protein expression between the target cell/tissue and the other cell(s)/tissue(s) will be. Conversely, the more such replacements, and/or the greater the difference in translational efficiencies of the synonymous between the different cells/tissues, the greater the difference in protein expression between the target cell/tissue and the other cell(s)/tissue(s) will be. The inventors have found in this respect that a protein can be expressed from a synthetic polynucleotide in a target cell/tissue at levels greater than 10,000-fold over those expressed in another cell/tissue.

In contrast to differential protein expression between different cells/tissues, it will be appreciated that a synthetic polynucleotide may be tailored with synonymous codons such that expression of a protein in a target cell is enhanced. In this regard, the difference in level of protein expressed in the target cell/tissue from a synthetic polynucleotide relative to that expressed from a parent polynucleotide depends on the percentage of existing codons

replaced by synonymous codons, and the difference in translational efficiencies between the existing codons and the synonymous codons in the target cell/tissue. Put another way, the fewer such replacements, and/or the smaller the difference in translational efficiencies between the synonymous and existing codons, the smaller the difference in protein expression between the synthetic polynucleotide and parent polynucleotide will be. Conversely, the more such replacements, and/or the greater the difference in translational efficiencies between the synonymous and existing codons, the greater the difference in protein expression between the synthetic polynucleotide and parent polynucleotide will be. The inventors have found in this respect that a protein can be expressed from a synthetic polynucleotide in a target cell/tissue at levels greater than 10,000-fold than from a parent polynucleotide.

Preferably, the at least one existing codon and the synonymous codon are selected such that said protein is expressed from said synthetic polynucleotide in said target cell or tissue at a level which is at least 110%, preferably at least 200%, more preferably at least 500%, and most preferably at least 1000%, of that expressed from said parent polynucleotide in said target cell or tissue.

In a preferred embodiment, the synonymous codon is a codon which has a higher translational efficiency in the target cell or tissue relative to a precursor cell or tissue of the target cell or tissue.

In an alternate embodiment, the synonymous codon is a codon which has a higher translational efficiency in the target cell or tissue relative to a cell or tissue derived from said target cell or tissue.

The two codons can be selected by measuring translational efficiencies of different codons in the target cell or tissue relative to the one or more other cells or tissues and identifying the at least one existing codon and the synonymous codon based on this measurement.

3. Methods of determining codon translational efficiency

3.1. Expressing a synthetic construct comprising a tandem repeat of identical codons fused in frame to a reporter polynucleotide

A major aspect of the present invention is based, at least in part, on the discovery that different but synonymous stretches of identical codons fused respectively in frame with a reporter polynucleotide can give rise to different levels of reporter protein expressed within a given cell type. Not wishing to be bound by any particular theory, it is believed that a tandem series of identical codons causes a ribosome to pause during translation if the iso-tRNA corresponding to the identical codons is limiting. In this regard, it is known that ribosomal pausing leads to a failure to complete a nascent polypeptide chain and an uncoupling of transcription and translation. Accordingly, the levels of reporter protein expressed in the different cells or tissues are sensitive to the intracellular abundance of the iso-tRNA species corresponding to the identical codons and therefore provide a direct correlation of a cell's or tissue's preference for translating a given codon. This means, for example, that if the levels of the reporter protein obtained in a cell or tissue type to which a synthetic construct having a tandem series of identical first codons is provided are lower than the levels expressed in the same cell or tissue type to which a different synthetic construct having a tandem series of identical

second codons is provided (*i.e.*, wherein the first codons are different from, but synonymous with, the second codons), then it can be deduced that the cell or tissue has a higher preference for the second codon relative to the first codon with respect to translation. Put another way, the second codon has a higher translational efficiency compared to the first codon in the cell or tissue type.

With regard to differential protein expression between different cell or tissue types, it will be appreciated that if the levels of the reporter protein obtained in a target cell or tissue type to which a synthetic construct having a tandem series of identical codons is provided are lower than the levels expressed in another cell or tissue type to which the same synthetic construct is provided, then it can be deduced that the target cell or tissue has a higher preference for the codon relative to the other cell or tissue with respect to translation. Put another way, the codon has a higher translational efficiency in the target cell or tissue compared to the other cell or tissue type.

Suitably, the tandem repeat comprises at least three identical codons. Preferably, the tandem repeat comprises four identical codons, more preferably five or seven identical codons and most preferably six identical codons.

The tandem repeat can be fused at a location adjacent to, or within, the reporter polynucleotide. The location is preferably selected such that the tandem repeat interferes with translation of at least a detectable portion of the reporter protein such that expression of the protein can be detected or assessed. Preferably, the tandem repeat is located immediately upstream (translationally) from the reporter polynucleotide.

It is of course possible that a tandem repeat of identical amino acid residues (*e.g.*, an oligo-proline repeat) can render the reporter protein unstable. Typically, protein instability is detected when expression of the reporter gene is not detectable with any choice of isoaccepting codon specific for the amino acid corresponding to the tandem repeat. The inventors have found in this regard that protein instability can be alleviated by use of at least one spacer codon within the tandem repeat of identical codons, wherein the spacer codon encodes a neutral amino acid.

The at least one spacer codon can be placed adjacent to, or interposed between, some or all of the identical codons corresponding to the tandem repeat. For example, a suitable interposition for a *penta*-repeat of identical codons can be selected from the group consisting of: (a) I-S-I-S-I-S-I-S-I-S; (b) S-I-S-I-S-I-S-I-S-I; (c) I-S-I-S-I-I-S-I; (d) I-S-I-I-S-I-S-I; (e) I-S-I-S-I-I-I; (f) I-I-S-I-S-I-I; (g) I-I-I-S-I-S-I; (h) I-S-I-I-S-I-I; (i) I-I-S-I-I-S-I; (j) I-S-I-I-I-S-I; (k) I-S-I-I-I-I; (l) I-I-S-I-I-I; (m) I-I-I-S-I-I; and (n) I-I-I-I-S-I, wherein I corresponds to an identical codon of a tandem repeat and S corresponds to a spacer codon.

Preferably, a spacer codon is efficiently translated in the cell or tissue type relative to other synonymous codons. This is important so that translation of the spacer codon is not rate limiting. The neutral amino acid includes, but is not restricted to, alanine and glycine.

The reporter polynucleotide can encode any suitable protein for which expression can be detected directly or indirectly such as by suitable assay. Suitable reporter polynucleotides include, but are not restricted to,

polynucleotides encoding β-galactosidase, firefly luciferase, alkaline phosphatase, chloramphenicol acetyltransferase (CAT), β-glucuronidase (GUS), herbicide resistance genes such as the bialophos resistance (BAR) gene that confers resistance to the herbicide BASTA, and green fluorescent protein (GFP). Assays for the activities associated with such proteins are known by those of skill in the art. Preferably, the reporter polynucleotide encodes GFP.

Persons of skill in the art will appreciate that reporter polynucleotides need not correspond to a full-length gene encoding a particular reporter protein. In this regard, the invention also contemplates reporter polynucleotide sub-sequences encoding desired portions of a parent reporter protein, wherein an activity or function of the parent protein is retained in said portions. A polynucleotide sub-sequence encodes a domain of the reporter protein having an activity associated therewith and preferably encodes at least 10, 20, 50, 100, 150, or 500 contiguous amino acid residues of the reporter protein.

The instant method is applicable to any suitable plant cell or tissue type. The cell or tissue type can be of any suitable lineage.

Suitable methods for isolating particular cells or tissues are known to those of skill in the art. For example, one can take advantage of one or more particular characteristics of a cell or tissue to specifically isolate the cell or tissue from a heterogeneous population. Such characteristics include, but are not limited to, anatomical location of a tissue, cell density, cell size, cell morphology, cellular metabolic activity, cell uptake of ions such as Ca^{2+}, K^{+}, and H^{+} ions, cell uptake of compounds such as stains, markers expressed on the cell surface, protein

fluorescence, and membrane potential. Suitable methods that can be used in this regard include surgical removal of tissue, flow cytometry techniques such as fluorescence-activated cell sorting (FACS), immunoaffinity separation (*e.g.*, magnetic bead separation such as Dynabead™ separation), density separation (*e.g.*, metrizamide, Percoll™, or Ficoll™ gradient centrifugation), and cell-type specific density separation.

In an alternate embodiment, progenitor cells or tissues can be used for initially introducing the synthetic construct. Any suitable progenitor cell or tissue can be used which gives rise to a particular plant cell or tissue of interest for which codon preference is to be ascertained. For example, a suitable progenitor of a plant includes, but is not restricted to, a meristematic cell and a callus tissue, respectively.

In another embodiment, the synthetic construct can be introduced first into a plant or part thereof before subsequent expression of the construct in a particular cell or tissue type of the plant.

The invention further provides for the analysis of codon translational efficiencies in a plant cell a synthetic construct comprising a reporter polynucleotide fused in frame with a tandem repeat of (e.g., 2, 3, 4, 5, 6, or 7 or more) identical codons, wherein said reporter polynucleotide encodes a reporter protein, and wherein said synthetic construct is operably linked to one or more regulatory polynucleotides.

The construction of the synthetic construct can be effected by any suitable technique. For example, *in vitro* mutagenesis methods can be employed, which are known to those of skill in the art. Suitable mutagenesis methods are

described for example in the relevant sections of Ausubel, *et al.* (*supra*) and of Sambrook, *et al.*, (*supra*) which are incorporated herein by reference. Alternatively, suitable methods for altering DNA are set forth, for example, in U.S. Patent Nos. 4,184,917, 4,321,365 and 4,351,901, which are incorporated herein by reference. Instead of *in vitro* mutagenesis, the synthetic polynucleotide can be synthesized *de novo* using readily available machinery. Sequential synthesis of DNA is described, for example, in U.S. Patent No 4,293,652, which is incorporated herein by reference. However, it should be noted that the present invention is not dependent on, and not directed to, any one particular technique for constructing the synthetic construct.

Regulatory polynucleotides which can be utilized to regulate expression of the synthetic polynucleotide include, but are not limited to, a promoter, an enhancer, and a transcription terminator. Such regulatory polynucleotides are known to those of skill in the art. The construct preferably comprises at least one promoter.

Regulatory polynucleotides which can be utilized to regulate expression of the synthetic construct include, but are not limited to, a promoter, an enhancer, and a transcriptional terminator. Such regulatory polynucleotides are known to those of skill in the art. Suitable promoters that can be utilized to induce expression of the polynucleotides of the invention include constitutive promoters and inducible promoters.

3.2. Determination of abundance of different tRNA species in and/or between different cells of a plant

The present invention contemplates any suitable method for determining the abundance of different iso-tRNA species in and/or between different cell or tissue types of a plant.

For example, such method can include isolating a particular cell or tissue from a plant, preparing an RNA extract from the cell or tissue which extract includes tRNA, and probing the extract with polynucleotides having different nucleic acid sequences, each being specific for a particular iso-tRNA to thereby determine the relative abundance of different iso-tRNAs in said cell or tissue. Preferably, this method is applied to two or more different cell or tissue types of the plant to determine the relative abundance of different iso-tRNAs between those cell or tissue types.

Suitable methods for isolating particular cells or tissues are known to those of skill in the art and are described, for example, in Section 3.1 above.

Any suitable method for isolating total RNA from a plant cell or tissue can be used. Typical procedures contemplated by the invention are described in CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (Ausubel, *et al.*, Eds.) (John Wiley & Sons, Inc. 1997), hereby incorporated by reference, at page 4.2.1 through page 4.2.7. Preferably, techniques which favor isolation of tRNA are employed as, for example, described in Brunngraber, E.F. (1962, *Biochem. Biophys. Res. Commun.* **8**: 1-3), which is hereby incorporated by reference.

The probing of an RNA extract is suitably effected with different oligonucleotide sequences each being specific for a particular iso-tRNA. Of course it will be appreciated that for a given plant, oligonucleotide sequences would need to be selected which hybridize specifically with particular iso-tRNA sequences expressed by the plant. Such selection is within the realm of one of ordinary skill in the art based on any known iso-tRNA sequence. Reference can be made in this regard to a compilation of tRNA sequences and sequences of tRNA genes described in Sprinzl *et al* (1996,

Nucleic Acids Res. **24**(1): 68-72; 1998, **26**(1): 148-53; 20 December, 1999 http://www.uni-bayreuth.de/departments/biochemie/trna/), the entire disclosures of which are incorporated herein by reference. For example, in the case of *Arabidopsis thaliana*, iso-tRNA sequences from which oligonucleotide sequences may be selected include, but are not limited to, those listed in TABLE 1 which sequences are incorporated herein by reference.

TABLE 1.

Nucleic acid sequences of various iso-tRNA species expressed by Arabidopsis thaliana

GenBank Accession #	Amino Acid	Codon Recognized	GenBank Accession #	Amino Acid	Codon Recognized
AB005780	Asp	GAC	D50933	Met	ATG
AB005781	Asp	GAC	D50934	Met	ATG
AB005779	Cys	TGC	M58320	Ser	TGA
D17336	Cys	TGC	L34745	Trp	TGG
D50935	Gln	CAA	L35907	Trp	TGG
AB005786	Glu	GAG	L35908	Trp	TGG
AB005782	Gly	GGC	L35909	Trp	TGG
AB005783	Gly	GGC	M35957	Tyr	AAT
AB005784	Gly	GGC	M35958	Tyr	TCA
AB005785	Gly	GGC	X54513	Val	GUU

Typically, the abundance of iso-tRNA species can be determined by blotting techniques that include a step whereby a sample RNA or tRNA extract is immobilized on a

matrix (preferably a synthetic membrane such as nitrocellulose), a hybridization step, and a detection step. Northern blotting can be used to identify an RNA sequence that is complementary to a polynucleotide probe. Alternatively, dot blotting and slot blotting can be used to identify complementary DNA/RNA or RNA/RNA nucleic acid sequences. Such techniques are known by those skilled in the art, and have been described, for example, in Ausubel, *et al* (*supra*) at pages 2.9.1 through 2.9.20.

According to such methods, a sample of tRNA immobilized on a matrix is hybridized under stringent conditions to a complementary polynucleotide (such as one having a sequence mentioned above) which is labeled, for example, radioactively, enzymatically or fluorochromatically.

While stringent washes are typically carried out at temperatures from about 42°C to 68°C, one skilled in the art will appreciate that other temperatures can be suitable for stringent conditions. Maximum hybridization typically occurs at about 20° to 25° below the T_m for formation of a DNA-DNA hybrid. It is known in the art that the T_m is the melting temperature, or temperature at which two complementary polynucleotides dissociate. Methods for estimating T_m are known in the art (see, e.g., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY *supra* at page 2.10.8). Maximum hybridization typically occurs at about 10° to 15° below the T_m for a DNA-RNA hybrid.

Other stringent conditions are known in the art. A skilled artisan will recognize that various factors can be manipulated to optimize the specificity of the hybridization. Optimization of the stringency of the final washes can serve to ensure a high degree of hybridization.

Methods for detecting labeled polynucleotides hybridized to an immobilized polynucleotide are known to practitioners in the art. Such methods include autoradiography, chemiluminescent, fluorescent and colorimetric detection.

Advantageously, the relative abundance of an iso-tRNA between two or more plant cells or tissues can be determined by comparing the respective levels of binding of a labeled polynucleotide specific for the iso-tRNA to equivalent amounts of immobilized RNA obtained from the two or more cells or tissues. Similar comparisons are suitably carried out to determine the respective relative abundance of other iso-tRNAs between the two or more cells or tissues. One of ordinary skill in the art will thereby be able to determine a relative tRNA abundance table (see for example TABLE 2 of International Application No. PCT/AU98/00530 the entire contents of which are hereby incorporated by reference) for different cells or tissues of a plant. From such comparisons, one or more synonymous codons can be selected such that the or each synonymous codon corresponds to an iso-tRNA which is in higher abundance in the target cell or tissue relative to other cells or tissues of the plant.

In the present embodiment, a synonymous codon is preferably selected such that its corresponding iso-tRNA is present in the target cell or tissue at a level which is at least 110%, preferably at least 200%, more preferably at least 500%, and most preferably at least 1000%, of that present in one or more other cells or tissues of the plant.

3.3. Analysis of codon usage

Alternatively, synonymous codons can be selected by analyzing the frequency at which codons are used by genes expressed in (i) particular cells or tissues of a plant,

(ii) substantially all cells or tissues of the plant, or (iii) an organism which can infect particular cells or tissues of the plant.

Codon frequency tables as well as suitable methods for determining frequency of codon usage in an organism are described, for example, in an article by Sharp *et al* (1988, *Nucleic Acids Res.* **16** 8207-8211), which is incorporated herein by reference. Reference also can be made to an article by Nakamura *et al* (2000, *Nucleic Acids Res* **28**(1): 292, incorporated herein by reference), which presents the frequency of each of the 257 468 complete protein-coding sequences (CDSs) compiled from the taxonomic divisions of the GenBank DNA sequence database. The sum of the codons used by 8792 organisms is also calculated. The data files relating to this article can be obtained from the anonymous ftp sites of DDBJ, Kazusa and EBI. A list of the codon usage of genes and the sum of the codons used by each organism can, for example, be obtained through the web site http://www.kazusa.or.jp/ codon/.

The relative level of gene expression (*e.g.*, detectable protein expression vs. no substantial or detectable protein expression) can provide an indirect measure of the relative translational efficiencies of codons, the relative abundance of specific iso-tRNAs expressed, or both, in different cells or tissues. For example, a virus can be capable of propagating within a first cell or tissue (which can include a cell or tissue at a specific stage of differentiation) but can be substantially incapable of propagating in a second cell or tissue (which can include a cell or tissue at another stage of differentiation). Comparison of the pattern of codon usage by genes of the virus with the pattern of codon usage by genes expressed in the second cell or tissue can thus provide indirectly a set of codons that

have high translational efficiencies and a set of codons that have low translational efficiencies in the first cell or tissue relative to the second cell or tissue. Simultaneously, the above comparison can also provide indirectly a set of codons that that have higher translational efficiencies and a set of codons that have low translational efficiencies in the second cell or tissue relative to the first cell or tissue.

From the foregoing, a synonymous codon according to the invention can correspond to a codon selected from the group consisting of (1) a codon used at relatively high frequency by genes, preferably highly expressed genes, of a target cell or tissue, (2) a codon used at relatively high frequency by genes, preferably highly expressed genes, of the plant, (3) a codon used at relatively low frequency by genes of one or more other cells or tissues, and (4) a codon used at relatively low frequency by genes of another organism.

By contrast, an existing codon according to the invention can correspond to a codon selected from the group consisting of (a) a codon used at relatively high frequency by genes, preferably highly expressed genes, of one or more other cells or tissues, (b) a codon used at relatively low frequency by genes of a target cell or tissue, (c) a codon used at relatively low frequency by genes of the plant, and (d) a codon used at relatively high frequency by genes of another organism.

Preferably, the genes from which codon frequency data are obtained do not relate to mitochondrial genes.

Suitably, a highly expressed gene according to the invention encodes a protein that is expressed at high levels, and preferably specifically (*i.e.*, substantially

only, *e.g.*, at a level at least about 100-fold greater than in other cells or tissues), in the target cell/tissue.

4. Construction of synthetic polynucleotides

The step of replacing synonymous codons for existing codons can be effected by any suitable technique. For example, *in vitro* mutagenesis methods can be employed which are known to those of skill in the art and include those described in Section 3.1 above.

The parent polynucleotide is preferably a natural gene. However, it is possible that the parent polynucleotide encodes a protein that is not naturally-occurring but has been engineered using recombinant techniques.

The parent polynucleotide need not be obtained from the plant but can be obtained from any suitable source, such as from a eukaryotic or prokaryotic organism. For example, the parent polynucleotide can be obtained from another plant or an animal. Broadly, the parent polynucleotide can be obtained from any eukaryotic or prokaryotic organism. In a preferred embodiment, the parent polynucleotide is obtained from a pathogenic organism. In such a case, a natural host of the pathogenic organism is preferably a plant. For example, the pathogenic organism can be a yeast, bacterium or virus.

For example, suitable proteins which may be used for selective expression in accordance with the invention include, but are not limited to *Bacillus thuringiensis* (Bt) proteins as for example described by Hill *et al.* (1995, *Euphytica* **85**(1-3):119-123, incorporated herein by reference) and Nayak *et al.* (1997, *Proc. Natl. Acad. Sci. USA* **94**(6):2111-2116, incorporated herein by reference), sunflower seed albumin as for example described by Tabe *et al.* (1993, *Genetica* **90**:181-200, incorporated herein by

reference), herbicide resistance proteins such as the BAR protein as for example described by Thompson *et al.* (1987, *EMBO J.* **6**:2519-2524, incorporated herein by reference).

The invention also contemplates synthetic polynucleotides encoding one or more desired portions of the protein to be expressed. A polynucleotide encodes a domain of the protein having a function associated therewith, or which is otherwise detectable, and preferably encodes at least 10, 20, 50, 100, 150, or 500 contiguous amino acid residues of the protein.

Regulatory polynucleotides which can be utilized to regulate expression from the synthetic polynucleotide include, but are not limited to, a promoter, an enhancer, and a transcriptional terminator. Such regulatory polynucleotides are known to those of skill in the art.

Suitable promoters which may be utilized to induce expression of the polynucleotides of the invention include constitutive promoters and inducible promoters. A particularly preferred promoter for dicotyledons which may be used to induce such expression is the Cauliflower Mosaic Virus (CaMV) 35S promoter. However, it will be appreciated that for monocotyledons, promoters including the ubiquitin promoter (p_{Ubi}) and the Emu promoter (p_{Emu}) may be employed.

Any suitable transcriptional terminator may be used which effects termination of transcription of polynucleotide. Preferably, the nopaline synthase (NOS) terminator, as for example disclosed in United States Patent Specification No. US 5,034,322, is used as the transcription terminator.

4.1. Expression vectors

Synthetic polynucleotides according to the invention can be operably linked to one or more regulatory polynucleotides in the form of an expression vector. The expression vector preferably contains an element(s) that permits stable integration of the vector into the host cell genome or autonomous replication of the vector in the cell independent of the genome of the cell. The vector may be integrated into the host cell genome when introduced into a host cell. For integration, the vector may rely on the foreign or endogenous DNA sequence or any other element of the vector for stable integration of the vector into the genome by homologous recombination. Alternatively, the vector may contain additional nucleic acid sequences for directing integration by homologous recombination into the genome of the host cell. The additional nucleic acid sequences enable the vector to be integrated into the host cell genome at a precise location in the chromosome. To increase the likelihood of integration at a precise location, the integrational elements should preferably contain a sufficient number of nucleic acids, such as 100 to 1,500 base pairs, preferably 400 to 1,500 base pairs, and most preferably 800 to 1,500 base pairs, which are highly homologous with the corresponding target sequence to enhance the probability of homologous recombination. The integrational elements may be any sequence that is homologous with the target sequence in the genome of the host cell. Furthermore, the integrational elements may be non-encoding or encoding nucleic acid sequences.

For autonomous replication, the vector may further comprise an origin of replication enabling the vector to replicate autonomously in the host cell in question. Examples of bacterial origins of replication are the origins

of replication of plasmids pBR322, pUC19, pACYC177, and pACYC184 permitting replication in *E. coli*, and pUB110, pE194, pTA1060, and pAM.beta.1 permitting replication in *Bacillus*. The origin of replication may be one having a mutation to make its function temperature-sensitive in a *Bacillus* cell (see, *e.g.*, Ehrlich, 1978, *Proc. Natl. Acad. Sci. USA* **75**:1433).

4.2. Marker genes

To facilitate identification of transformants, the synthetic construct desirably comprises a selectable or screenable marker gene as, or in addition to, the synthetic polynucleotide. The actual choice of a marker is not crucial as long as it is functional (*i.e.*, selective) in combination with the plant cells of choice. The marker gene and the foreign or endogenous DNA sequence of interest do not have to be linked, since co-transformation of unlinked genes as, for example, described in U.S. Pat. No. 4,399,216 is also an efficient process in plant transformation.

Included within the terms selectable or screenable marker genes are genes that encode a "secretable marker" whose secretion can be detected as a means of identifying or selecting for transformed cells. Examples include markers that encode a secretable antigen that can be identified by antibody interaction, or secretable enzymes that can be detected by their catalytic activity. Secretable proteins include, but are not restricted to, proteins that are inserted or trapped in the cell wall (*e.g.*, proteins that include a leader sequence such as that found in the expression unit of extensin or tobacco PR-S); small, diffusible proteins detectable, *e.g.* by ELISA; and small active enzymes detectable in extracellular solution (*e.g.*, α-amylase, β-lactamase, phosphinothricin acetyltransferase).

4.2.1. Selectable markers

A selectable marker is a gene the product of which provides for biocide resistance, resistance to heavy metals, prototrophy to auxotrophs, and the like. Examples of bacterial selectable markers are the *dal* genes from *Bacillus subtilis* or *Bacillus licheniformis*, or markers that confer antibiotic resistance such as ampicillin, kanamycin, erythromycin, chloramphenicol or tetracycline resistance. Exemplary selectable markers for selection of plant transformants include, but are not limited to, a *hyg* gene which encodes hygromycin B resistance; a neomycin phosphotransferase (*neo*) gene conferring resistance to kanamycin, paromomycin, G418 and the like as, for example, described by Potrykus *et al.* (1985, *Mol. Gen. Genet.*, **199**:183); a glutathione-S-transferase gene from rat liver conferring resistance to glutathione derived herbicides as, for example, described in EP-A 256 223; a glutamine synthetase gene conferring upon overexpression resistance to glutamine synthetase inhibitors such as phosphinothricin as, for example, described WO87/05327, an acetyl transferase gene from *Streptomyces viridochromogenes* conferring resistance to the selective agent phosphinothricin as, for example, described in EP-A 275 957, a gene encoding a 5-enolshikimate-3-phosphate synthase (EPSPS) conferring tolerance to N-phosphonomethylglycine as, for example, described by Hinchee *et al.* (1988, *Biotech.*, **6**:915), a *bar* gene conferring resistance against bialophos as, for example, described in WO91/02071; a nitrilase gene such as *bxn* from *Klebsiella ozaenae* which confers resistance to bromoxynil (Stalker *et al.*, 1988, *Science*, **242**:419); a methotrexate-resistant DHFR gene (Thillet *et al.*, 1988, *J. Biol. Chem.*, **263**:12500); a mutant acetolactate synthase gene (ALS), which confers resistance to imidazolinone,

sulfonylurea or other ALS-inhibiting chemicals (European Patent Application 154,204, 1985); a mutated anthranilate synthase gene that confers resistance to 5-methyl tryptophan; or a dalapon dehalogenase gene that confers resistance to the herbicide.

4.2.2. Screenable markers

A screenable marker is a gene the product of which provides for a detectable trait. Preferred screenable markers include, but are not limited to, a *uidA* gene encoding a β-glucuronidase (GUS) enzyme for which various chromogenic substrates are known; a β-galactosidase gene encoding an enzyme for which chromogenic substrates are known; an aequorin gene (Prasher *et al.*, 1985, *Biochem. Biophys. Res. Comm.*, **126**:1259), which may be employed in calcium-sensitive bioluminescence detection; a green fluorescent protein gene (Niedz *et al.*, 1995 *Plant Cell Reports*, **14**:403); a luciferase (*luc*) gene (Ow *et al.*, 1986, *Science*, **234**:856), which allows for bioluminescence detection; a β-lactamase gene (Sutcliffe, 1978, *Proc. Natl. Acad. Sci. USA* **75**:3737), which encodes an enzyme for which various chromogenic substrates are known (*e.g.*, PADAC, a chromogenic cephalosporin); an R-locus gene, encoding a product that regulates the production of anthocyanin pigments (red color) in plant tissues (Dellaporta *et al.*, 1988, in *Chromosome Structure and Function*, pp. 263-282); an α-amylase gene (Ikuta *et al.*, 1990, *Biotech.*, **8**:241); a tyrosinase gene (Katz *et al.*, 1983, *J. Gen. Microbiol.*, **129**:2703) which encodes an enzyme capable of oxidizing tyrosine to DOPA and dopaquinone which in turn condenses to form the easily detectable compound melanin; or a *xylE* gene (Zukowsky *et al.*, 1983, *Proc. Natl. Acad. Sci. USA* **80**:1101),

which encodes a catechol dioxygenase that can convert chromogenic catechols.

5. Introduction of the synthetic polynucleotide in a plant cell

The step of introducing the synthetic polynucleotide into a target cell or tissue will differ depending on the intended use and or species, and may involve infection by *Agrobacterium tumefaciens* or *A rhizogenes*, electroporation, micro-projectile bombardment or protoplast fusion.

A number of techniques are available for the introduction of DNA into a plant host cell. There are many plant transformation techniques well known to workers in the art, and new techniques are continually becoming known. The particular choice of a transformation technology will be determined by its efficiency to transform certain plant species as well as the experience and preference of the person practising the invention with a particular methodology of choice. It will be apparent to the skilled person that the particular choice of a transformation system to introduce a the synthetic polynucleotides of the invention into plant cells is not essential to or a limitation of the invention, provided it achieves an acceptable level of nucleic acid transfer. Guidance in the practical implementation of transformation systems for plant improvement is provided, for example, by Birch (1997, *Annu. Rev. Plant Physiol. Plant Molec. Biol.* **48**: 297-326), which is incorporated herein by reference.

In principle both dicotyledonous and monocotyledonous plants that are amenable to transformation, can be modified by introducing a synthetic polynucleotide according to the invention into a recipient cell and growing a new plant that harbors and expresses the synthetic polynucleotide.

Introduction and expression of synthetic polynucleotides in dicotyledonous (broadleaved) plants such as tobacco, potato and alfalfa has been shown to be possible using the T-DNA of the tumor-inducing (Ti) plasmid of *Agrobacterium tumefaciens* (See, for example, Umbeck, U.S. Patent No. 5,004,863, and International application PCT/US93/02480). A construct of the invention may be introduced into a plant cell utilizing *A tumefaciens* containing the Ti plasmid. In using an *A. tumefaciens* culture as a transformation vehicle, it is most advantageous to use a non-oncogenic strain of the *Agrobacterium* as the vector carrier so that normal non-oncogenic differentiation of the transformed tissues is possible. It is preferred that the *Agrobacterium* harbors a binary Ti plasmid system. Such a binary system comprises (1) a first Ti plasmid having a virulence region essential for the introduction of transfer DNA (T-DNA) into plants, and (2) a chimeric plasmid. The chimeric plasmid contains at least one border region of the T-DNA region of a wild-type Ti plasmid flanking the nucleic acid to be transferred. Binary Ti plasmid systems have been shown effective to transform plant cells as, for example, described by De Framond (1983, *Biotechnology*, **1**:262) and Hoekema *et al.* (1983, *Nature*, **303**:179). Such a binary system is preferred because it does not require integration into Ti plasmid in *Agrobacterium*.

Methods involving the use of *Agrobacterium* include, but are not limited to: (a) co-cultivation of *Agrobacterium* with cultured isolated protoplasts; (b) transformation of plant cells or tissues with *Agrobacterium*; or (c) transformation of seeds, apices or meristems with *Agrobacterium*.

Recently, rice and corn, which are monocots, have been shown to be susceptible to transformation by *Agrobacterium* as well. However, many other important monocot crop plants,

including oats, sorghum, millet, and rye, have not yet been successfully transformed using *Agrobacterium*-mediated transformation. The Ti plasmid, however, may be manipulated in the future to act as a vector for these other monocot plants. Additionally, using the Ti plasmid as a model system, it may be possible to artificially construct transformation vectors for these plants. Ti plasmids might also be introduced into monocot plants by artificial methods such as microinjection, or fusion between monocot protoplasts and bacterial spheroplasts containing the T-region, which can then be integrated into the plant nuclear DNA.

In addition, gene transfer can be accomplished by *in situ* transformation by *Agrobacterium*, as described by Bechtold *et al.* (1993, *C.R. Acad. Sci. Paris*, **316**:1194). This approach is based on the vacuum infiltration of a suspension of *Agrobacterium* cells.

Alternatively, foreign or chimeric nucleic acids may be introduced using root-inducing (Ri) plasmids of *Agrobacterium* as vectors.

Cauliflower mosaic virus (CaMV) may also be used as a vector for introducing of exogenous nucleic acids into plant cells (U.S. Pat. No. 4,407,956). CaMV viral DNA genome is inserted into a parent bacterial plasmid creating a recombinant DNA molecule that can be propagated in bacteria. After cloning, the recombinant plasmid again may be cloned and further modified by introduction of the desired nucleic acid sequence. The modified viral portion of the recombinant plasmid is then excised from the parent bacterial plasmid, and used to inoculate the plant cells or plants.

Synthetic polynucleotides according to the invention can also be introduced into plant cells by electroporation as, for example, described by Fromm *et al.* (1985, *Proc. Natl. Acad. Sci., U.S.A*, **82**:5824) and Shimamoto *et al.* (1989, *Nature* **338**:274-276). In this technique, plant protoplasts are electroporated in the presence of vectors or nucleic acids containing the relevant nucleic acid sequences. Electrical impulses of high field strength reversibly permeabilize membranes allowing the introduction of nucleic acids. Electroporated plant protoplasts reform the cell wall, divide and form a plant callus.

Another method for introducing synthetic polynucleotides into a plant cell is high velocity ballistic penetration by small particles (also known as particle bombardment or microprojectile bombardment) with the nucleic acid to be introduced contained either within the matrix of small beads or particles, or on the surface thereof as, for example described by Klein *et al.* (1987, *Nature* **327**:70). Although, typically only a single introduction of a new nucleic acid sequence is required, this method particularly provides for multiple introductions.

Alternatively, the synthetic polynucleotides can be introduced into a plant cell by contacting the plant cell using mechanical or chemical means. For example, a nucleic acid can be mechanically transferred by microinjection directly into plant cells by use of micropipettes. Alternatively, a nucleic acid may be transferred into the plant cell by using polyethylene glycol which forms a precipitation complex with genetic material that is taken up by the cell.

There are a variety of methods known currently for transformation of monocotyledonous plants. Presently, preferred methods for transformation of monocots are

microprojectile bombardment of explants or suspension cells, and direct DNA uptake or electroporation as, for example, described by Shimamoto *et al.* (1989, *supra*). Transgenic maize plants have been obtained by introducing the *Streptomyces hygroscopicus bar* gene into embryogenic cells of a maize suspension culture by microprojectile bombardment (Gordon-Kamm, 1990, *Plant Cell*, **2**:603-618). The introduction of genetic material into aleurone protoplasts of other monocotyledonous crops such as wheat and barley has been reported (Lee, 1989, *Plant Mol. Biol.* **13**:21-30). Wheat plants have been regenerated from embryogenic suspension culture by selecting only the aged compact and nodular embryogenic callus tissues for the establishment of the embryogenic suspension cultures (Vasil, 1990, *Bio/Technol.* **8**:429-434). The combination with transformation systems for these crops enables the application of the present invention to monocots. These methods may also be applied for the transformation and regeneration of dicots. Transgenic sugarcane plants have been regenerated from embryogenic callus as, for example, described by Bower et al. (1996, *Molecular Breeding* **2**:239-249).

Alternatively, a combination of different techniques may be employed to enhance the efficiency of the transformation process, *e.g.*, bombardment with *Agrobacterium* coated microparticles (EP-A-486234) or microprojectile bombardment to induce wounding followed by co-cultivation with *Agrobacterium* (EP-A-486233).

The literature articles referred to above are all incorporated herein by reference.

6. Production and characterisation of differentiated transgenic plants

6.1. Regeneration

The methods used to regenerate transformed cells into differentiated plants are not critical to this invention, and any method suitable for a target plant can be employed. Normally, a plant cell is regenerated to obtain a whole plant from a transformation process.

Regeneration from protoplasts varies from species to species of plants, but generally a suspension of protoplasts is first made. In certain species, embryo formation can then be induced from the protoplast suspension, to the stage of ripening and germination as natural embryos. The culture media will generally contain various amino acids and hormones, necessary for growth and regeneration. Examples of hormones utilized include auxin and cytokinins. It is sometimes advantageous to add glutamic acid and proline to the medium, especially for such species as corn and alfalfa. Efficient regeneration will depend on the medium, on the genotype, and on the history of the culture. If these variables are controlled, regeneration is reproducible. Regeneration also occurs from plant callus, explants, organs or parts. Transformation can be performed in the context of organ or plant part regeneration as, for example, described in *Methods in Enzymology*, Vol. 118 and Klee *et al.* (1987, *Annual Review of Plant Physiology*, **38**:467), which are incorporated herein by reference. Utilizing the leaf disk-transformation-regeneration method of Horsch *et al.* (1985, *Science*, **227**:1229, incorporated herein by reference), disks are cultured on selective media, followed by shoot formation in about 2-4 weeks. Shoots that develop are excised from calli and transplanted to appropriate root-inducing

selective medium. Rooted plantlets are transplanted to soil as soon as possible after roots appear. The plantlets can be repotted as required, until reaching maturity.

In vegetatively propagated crops, the mature transgenic plants are propagated by the taking of cuttings or by tissue culture techniques to produce multiple identical plants. Selection of desirable transgenotes is made and new varieties are obtained and propagated vegetatively for commercial use.

In seed propagated crops, the mature transgenic plants can be self-crossed to produce a homozygous inbred plant. The inbred plant produces seed containing the newly introduced foreign gene(s). These seeds can be grown to produce plants that would produce the selected phenotype, *e.g.*, early flowering.

Parts obtained from the regenerated plant, such as flowers, seeds, leaves, branches, fruit, and the like are included in the invention, provided that these parts comprise cells that have been transformed as described. Progeny and variants, and mutants of the regenerated plants are also included within the scope of the invention, provided that these parts comprise the introduced synthetic polynucleotides of the invention.

It will be appreciated that the literature describes numerous techniques for regenerating specific plant types and more are continually becoming known. Those of ordinary skill in the art can refer to the literature for details and select suitable techniques without undue experimentation.

6.2. Characterisation

To confirm the presence of the synthetic polynucleotide in the regenerating plants, a variety of assays may be

performed. Such assays include, for example, "molecular biological" assays well known to those of skill in the art, such as Southern and Northern blotting and PCR; a protein expressed by the synthetic polynucleotide may be analysed by high performance liquid chromatography or ELISA (*e.g.*, nptII) as is well known in the art.

7. *Applications*

In one embodiment of the present invention, the target cell is suitably a cell or tissue of a plant leaf. Advantageously, the protein which is desired to be selectively expressed in the leaf cell or tissue is not expressible in a cell or tissue of the plant's root from a parent polynucleotide at a level sufficient to effect a particular function associated with said protein. In this embodiment, the step of replacing at least one existing codon with a synonymous codon is characterized in that the synonymous codon has a higher translational efficiency in the leaf cell or tissue compared to the root cell or tissue. Accordingly, a synthetic polynucleotide is produced having altered translational kinetics compared to the parent polynucleotide wherein the protein is expressible in the leaf cell or tissue at a level sufficient to effect a particular function associated with said protein, but wherein the protein is not expressible in the root cell or tissue at a level sufficient to effect said function.

The above embodiment may be utilized advantageously for conferring herbicide resistance in leaves of a plant, but not in roots of the plant where overexpression of a herbicide resistance gene may adversely affect plant growth. In such a case, a suitable protein may include the BAR protein which confers resistance to the herbicide BASTA.

8. Expressing a protein in a target cell or tissue by in vivo expression of iso-tRNAs in the target cell or tissue

The invention also extends to a method wherein a protein can be selectively expressed in a target cell of a plant by introducing into the cell an auxiliary polynucleotide capable of expressing in the target cell one or more isoaccepting transfer RNAs which are not normally expressed in relatively high abundance in the target cell but which are rate-limiting for expression of the protein from a parent polynucleotide to a level sufficient for effecting a function associated with the protein. In this embodiment, introduction of the auxiliary polynucleotide sequence in the target cell changes the translational kinetics of the parent polynucleotide such that said protein is expressed at a level sufficient to effect a function associated with the protein.

The step of introducing the auxiliary polynucleotide sequence into the target cell or a tissue comprising a plurality of these cells can be effected by any suitable means. For example, analogous methodologies for introduction of the synthetic polynucleotide referred to above can be employed for delivery of the auxiliary polynucleotide into said target cell.

In practice, the choice of iso-tRNA supplied to a target cell using this method depends on whether the target cell in which protein expression is desired has a low or high abundance of that iso-tRNA and on whether the parent polynucleotide comprises codons corresponding to that iso-tRNA species. Thus, an iso-tRNA is supplied to the target cell by the auxiliary polynucleotide when that iso-tRNA is in relatively low abundance in the target cell and when

parent polynucleotide comprises codons corresponding to that iso-tRNA species.

The invention is further described with reference to the following non-limiting examples.

EXAMPLE 1

Construction of expression vectors for determining relative abundance of different iso-tRNAs in different cells or tissues of Arabidopsis thaliana.

An oligonucleotide (SEQ ID NO:23; sequence complementary to the termination codon for GFP, is underlined), and a suite of oligonucleotides (SEQ ID NO: 24, 25, 26, 27, 28, 29, 30, 31, 32, 33 and 34; the first codon of GFP, is underlined) will be synthesized and used for PCR amplification of pBIN m-gfp5-ER$^-$-RS$^-$ (SEQ ID NO:35) template with *Taq* DNA polymerase (Amplification parameters: 94°C/1 min; 55°C/1 min; 72°C/2 min for 30 cycles). The amplified fragments will have nucleic acid sequences as shown in SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, and 21. The deduced amino acid sequences encoded by these synthetic fragments are shown below the corresponding nucleic acid sequences as well as in SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20 and 22. In summary, the synthetic fragments contain an artificial start codon followed by a tandem repeat of six identical codons specific for a given iso-tRNA species. The tandem repeat immediately precedes the second codon of the *gfp* gene. The synthetic fragments by SEQ ID NO and tandem repeat encoded thereby are presented in the TABLE 2.

TABLE 2

Synthetic fragments and tandem repeats encoded thereby.

SEQ ID NO	Tandem repeat	SEQ ID NO	Tandem repeat
1	control	13	Leu (CTT) x 6
3	His (CAT) x 6	15	Leu (TTA) x 6
5	His (CAC) x 6	17	Leu (TTG) x 6
7	Leu (CTA) x 6	19	Lys (AAA) x 6
9	Leu (CTC) x 6	21	Lys (AAG) x 6
11	Leu (CTG) x 6		

The amplified fragments will be cloned using either the *Hind*III or *Bam*HI sites at the 5′ end and either the *Pst*I or *Eco*RI sites at the 3′ end of the pAOV2 polylinker (Mylne and Botella, 1999, *Plant Mol. Biol. Res.*, in press; the manuscript relating thereto is annexed hereto as Annexure A) for expression in *Escherichia coli* and for plant transformation experiments.

EXAMPLE 2

Agrobacterium Tri-Parental Mating

Tri-parental mating uses natural conjugation between bacteria to transfer a plasmid from *E. coli* (donor) to *Agrobacterium* (recipient) with the conjugal assistance from (helper) *E. coli* (HB101/pRK2013). Successful conjugation is selected for by combinatorial selection with antibiotics.

Agrobacterium media:

There are a number of possible media that can be used for *Agrobacterium* growth including (Minimal A, LB, YEP, YEB, YM etc). Some of these media provide additional selection for *Agrobacterium* by using sucrose as the primary carbon source rather than glucose. *E. coli* cannot metabolize

sucrose well. The following procedures all involve use of LB for *Agrobacterium* growth.

Procedure:

Plates are made for recipient *Agrobacterium* (LBA4404: LB-Rif^{50}/Strep^{25} or GV3101: LB-Rif^{50}/Gen^{25} - Rif^{50} alone); for donor (e.g. LB-Tet^{10} for pAOV2 - plates depend on the vector used) and for helper (LB-Kan^{50}); as well an LB (only) plate for the mating. Plates are also made that contain antibiotics for the vector to be used as well as for *Agrobacterium* (e.g. LB-Rif^{50}/Tet^{2}).

The bacteria are grown on the plates (start *Agrobacterium* first) until good sized individual colonies are formed. A loop is used to take a colony from each plate and these colonies are mixed together in the middle of the LB plate. This plate is then incubated at 28°C for 24 hours.

A loop of the mix is then plated out (16 streak) onto a combinatorial selection plate. This plate is then placed in a 28°C incubator for 2-3 days after which isolated individual colonies are picked for culture (at this stage, some of the bacteria may be used to screen for recombinant *Agrobacterium* by PCR) - Note it is important to "ease" the bacteria back onto selection (LB-Rif^{50}/Tet^{2}/Strep^{25} or Gen^{25} all at once after mating is too strong and no colonies will grow back from the mating).

A colony from the above plate is then exposed to stronger selection before subsequent growth and PCR (e.g. plates containing LB-Rif^{100}/Tet^{2}/Strep^{25})

Control plate:

If tri-parental mating is desired to be validated, a control plate may be run in tandem - i.e. LB plate with only

two of D,H or R (D - donor; H - helper; R - recipient). This plate is incubated at 28°C for 24 hours. A loop of the mixes is then plated out (16 streak) onto a combinatorial selection plate. And the plate is then placed at 28°C for 2-3 days -no colonies should be obtained (however, a few colonies may be obtained for the DR - because of a low rate of conjugation). This control validates that your helper indeed helps and that the combinatorial selection is working.

EXAMPLE 3

Screening Agrobacterium by PCR

A portion of a large (matchhead-size) *Agrobacterium* colony is used to inoculate selective media. 20 µl of the following PCR PreMix is then placed into a PCR tube.

PCR PreMix

dNTP's @ 10 mM	50 µL
Commercial 35S Forward Primer @10 µM	50 µL
Commercial NOS Reverse Primer @ 10 µM	50 µL
$MgCl_2$ @ 25 mM	150 µL
10X PCR Buffer II	250 µL
H_2O	1450 µl

A Gilson Pipetteman™ P_{10} tip is used to stab the same *Agrobacterium* colony from the plate and the stab of bacteria is mixed with 20 µL of the PreMix. Two microliters (2 Units) of Taq DNA polymerase , 3 µL of H_2O and 40 µL of mineral oil are then added to the mixture.

Mixtures are then amplified by PCR using the following parameters:

94°C	1 min	
94°C	30 sec	35 cycles
56°C	30 sec	
72°C	3 min	
72°C	15 min	
4°C	HOLD	

After completion, 10 µL of the PCR reaction is then subjected to agarose gel electrophoresis to confirm whether PCR products have been obtained.

EXAMPLE 4

Transformation of Arabidopsis by Vacuum Infiltration (in planta)

Vacuum infiltration involves putting adult Arabidopsis plants under a suspension of Agrobacterium in a vacuum chamber. A vacuum is drawn and the air spaces in the plant expand and bubble to the surface. When the pressure is returned, the Agrobacterium solution replaces the air spaces and the Agrobacterium are introduced into the inside of the plant. The Agrobacterium transform, among other cells, parts of the floral meristem and some of these cells will give rise to embryos and eventually, when planted, a whole transgenic plant.

Arabidopsis preparation

An 8cm diameter pot is filled with a moist soil mixture. The pot is covered with a 15 cm diameter (flyscreen) mesh circle, and an elastic band is used to secure it, keeping the mesh taught across the top, and making sure that the soil touches the mesh.

Mancozeb 2g/L (Yates Mancozeb Plus Garden Fungicide) is sprayed onto the soil and Arabidopsis (ecotype Columbia) - approximately 20 or so per pot - is sprinkled thereon.

The pots are placed into a Yates seedling chamber (which holds about 16 pots), and a lid is placed upon the containers. The containers are then placed in a 4°C cold room for 3-5 days for stratification (the cold snap synchronizes germination).

The plants are then placed under short day lights (*i.e.*, 8hr day) for 10-14 days (short day encourages vegetative growth). Seedlings may be thinned out if required. However, the plants should be kept well watered when young.

Pots are then placed in long day (i.e. 16 hr day) to induce bolting. The plants are then watered from below about every 2-3 days (plants require more water when they are filling their siliques). Bolts may be pruned back if necessary to encourage numerous unopened buds at time of infiltration (by clipping primary inflorescence at its base, and letting secondary inflorescences grow until they start to show open flowers).

The above method has been used successfully to transform mature plants and even non-bolted plants. However, it will be appreciated that plants should not be water stressed , otherwise their stomata will be shut and infiltration will not be efficient. Plants should be watered the evening before to ensure the stomata are open for the infiltration. It will be appreciated that healthy plants and correct infiltration media are important to a successful transformation.

Agrobacterium preparation

When the plants begin to bolt, Agrobacterium-pAOV2 constructs are plated onto LB-$Rif^{50}Tet^{2}$ at 28°C. A 5 mL LB-$Rif^{50}Tet^{2}$ culture is set up with fresh Agrobacterium from the plates and shaken at 28°C for 2 days. The culture is then added to a 100 mL Erlenmeyer flask containing 35 mL LB-$Rif^{50}Tet^{2}$ and this is shaken overnight before addition to a 1 L Erlenmeyer flask containing 400 mL LB-Tet^{2} (No Rif) and subsequent culturing at 28°C for about 24 hrs (or until Agrobacterium is in late log/early stationary phase)

Infiltration

The Agrobacterium culture (ideal OD_{600} is about 1.6-2.0)is then harvested by centrifugation in a GSA rotor (Sorvall) at 5000 rpm for 20 min at room temperature (however, infiltration still works if centrifugation is performed at 4°C).

The resulting cell pellet is resuspended in ~1 L infiltration media (see below). Some of the infiltration culture is poured into two 1 L containers which are then placed in the vacuum chamber. A skewer is placed through opposite holes in the bottom of two pots containing *A. thaliana* and the pots are inverted in the containers. Each pot should be suspended upside-down in a bottle with the skewer holding the pot a few centimeters from the bottom. In this regard, it is important to ensure that the entire bolts of the plants are covered by the culture (especially the flowers) - fungal infection may result if too much infiltration media enters the soil and rosette.

The vacuum chamber is then closed, before closing the air outlet of the chamber, and drawing a vacuum until

solution bubbles vigorously. The vacuum is then released as soon as possible.

The infiltration procedure is then repeated on the same plants.

Plants are placed on their sides in the containers. The top of the containers are then covered with a plastic wrap before placing the containers under long day light for one night. The next day the plastic wrap is removed and the plants are set upright.

Plants should not be watered until approximately 6 days after infiltration (although plants should be watered if they are looking wilted).

Plants are then grown until bolts start to yellow on some siliques. Seed catchers are then placed on the plants. For example, seed catchers may be made from soft-drink bottles by cutting off the bottom of the bottles and making a hole in each bottle near the neck (about where the curve begins to straighten out into the cylinder of the bottle).

When flowers terminate into siliques, watering of the plants is stopped and when the entire plant is dry, the seeds are harvested (seeds from green siliques contain germination inhibitors)

<u>Basta® selection of transformants</u>

Approximately 0.5 mL of seed is planted in a seedling tray containing moist soil, and this is covered with vermiculite dust and a chamber lid before placing at 4°C to synchronize germination (3-5 days).

The chambers are placed under long day light and at emergence, the plants are sprayed with 2000 mg/L (active component of Basta® - is glufosinate ammonium at a

concentration of 200g/L in Hoechst Horticulture Basta® U.N. No 2902).

The plants are then sprayed three days later and again three days after that. Transgenic plants will be very green and much larger than non-transgenic plants which should wither and die. Transgenic plants can be transferred later to individual pots if desired.

Routinely, approximately 75-100 transformants from five plants are isolated using this procedure.

Solution and media used for infiltration are presented in TABLE 3.

TABLE 3

Solutions and Media

Infiltration media	1 liter	2 liters	5 liters
0.5x MS salts	2.16 g	4.32 g	10.8 g
5% w/v Sucrose	50 g	100 g	250 g
1x B5 (or MS) vitamins	1 mL @ 1000X	2 mL @ 1000X	5 mL @ 1000X
0.25 g/L MES	0.25 g	0.5 g	1.25 g
pH to 5.7 (with KOH)			
After autoclaving add:			
10 µL/L of 1 mg/mL BAP			
50 µL/L Silwet L-77 (detergent)			

EXAMPLE 5

Confocal microscopy

Transgenic *Arabidopsis* seedlings are grown in sterile agar culture for 5 days, and are mounted in water under glass coverslips for microscopy. The specimens are examined using a Bio-Rad MRC-600 laser-scanning confocal microscope equipped with a krypton-argon laser and filter sets suitable for the detection of fluorescein and Texas red dyes (Bio-Rad K1yK2), and a Nikon 603 PlanApo numerical aperture 1.2 water-immersion objective. Dual-channel confocal images and video montages of seedlings are suitably composed using ADOBE PhotoShop.

The disclosure of every patent, patent application, and publication cited herein is hereby incorporated by reference in its entirety.

The present invention has been described in terms of particular embodiments found or proposed by the present inventors to comprise preferred modes for the practice of the invention. Those of skill in the art will appreciate that, in light of the present disclosure, numerous modifications and changes can be made in the particular embodiments exemplified without departing from the scope of the invention. All such modifications are intended to be included within the scope of the appended claims.

ANNEXURE A

Quick Sense and Antisense EST Constructs

Genetic Resources

Binary Vectors for Sense and Antisense Expression of *Arabidopsis* ESTs

JOSHUA MYLNE AND JOSÉ RAMÓN BOTELLA[1]

Plant Genetic Engineering Laboratory, Department of Botany, The University of Queensland, Brisbane, Queensland, 4072, Australia.

[1]Corresponding author. E-mail: J.Botella@botany.uq.edu.au; tel 61-7-3365 1128; fax 61-7-3365 1699
Abbreviations: ABRC, *Arabidopsis* biological resource center ; CaMV, cauliflower mosaic virus; EST, expressed sequence tag; NASC, Nottingham *Arabidopsis* stock centre; PCR, polymerase chain reaction

Abstract

Our laboratory is interested in devising methods to allow the identification of the putative functions of the vast pool of genes readily available for *Arabidopsis*. For this purpose we have constructed a set of binary vectors that will allow the quick production of transgenic *Arabidopsis* plants containing either sense or antisense copies of EST clones obtained from the PRL2 library (Newman et al., 1994). These vectors are based on the pSLJ series (Jones et al., 1992) containing the bialophos resistance (BAR) gene that confers resistance to the herbicide BASTA (Thompson et al., 1987). In addition, our vectors contain a 35S CaMV promoter-polylinker-*nos* terminator cassette that allows the direct cloning of *Arabidopsis* ESTs in either antisense (pAOV and pAOV2) or sense (pSOV and pSOV2) orientation. We also describe the construction of two additional vectors conferring BASTA resistance and containing the pBluescript polylinker in both orientations inserted between the 35S CaMV promoter and *nos* terminator (pKMB and pSMB).

Keywords

antisense constructs, *Arabidopsis*, binary vector, EST, herbicide resistance, transgenic, sense constructs.

Introduction

Thanks to the ongoing plant genome projects, the emphasis in molecular biology has shifted from cloning genes to establishing the function of available genes. Among the many methods available to study gene function, two of the most popular ones are the

production of loss-of-function and/or gain-of-function mutants. When a laboratory clones a new gene of unknown function, one of the first approaches which comes to mind is to knock out the gene and study the phenotype of the resulting mutant plants. Down-regulation or over-expression of a given gene is most frequently achieved by the production of transgenic plants carrying antisense or extra sense copies of the gene. Unfortunately not all plants are amenable to transformation and not all genes can be studied by this method.

The main considerations when planning to produce transgenic plants carrying either sense or antisense constructs are (a) the species to use, (b) the transformation method and (c) the selectable marker. *Arabidopsis* is the model system used by many laboratories due to the simplicity of its genome and availability of standard methods for transformation. Recent developments in *Arabidopsis* transformation by *in planta* vacuum infiltration (Bechtold et al., 1993) have avoided the need for tissue culture, providing an extremely easy method to obtain transgenic plants without the help of any specialized equipment. Furthermore, the availability of herbicide resistance genes (De Block et al., 1987; Lee et al., 1988) has superseded the need to use antibiotic resistance as selectable marker and sterile techniques for the selection of transformants.

Our vectors take advantage of all the above mentioned developments and allows direct cloning of ESTs in either sense or antisense orientations that can be combined with *in planta* transformation and direct selection of transgenic plants in seed trays by BASTA resistance.

Materials and Methods

Standard techniques were used for bacterial growth (*E.coli* DH10B) and DNA manipulation (Sambrook et al., 1989). Large amounts of high purity binary vector DNA required for cloning or sequencing were obtained as described by Jones *et al.* (1992).

Vector construction

All vectors described in this paper have been constructed using pSLJ75515 as backbone and their general outlay is shown in **Fig.** 1.

pAOV (Antisense Orientation Vector) (Fig. 2) was constructed by separately excising the 35S CaMV promoter and *nos* terminator from pBI121 (Bevan, 1984) and ligating them into the pUC18 multiple cloning site. This 35S-polylinker-*nos* cassette was cloned into pSLJ75515 which had previously had superfluous restriction enzyme sites removed.

pSOV (Sense Orientation Vector) (Fig. 2) was constructed by separately excising the 35S CaMV promoter and *nos* terminator from pBI121 and ligating them into pZLD. pZLD was obtained by removing the EST insert from a PRL2 clone with *Sal*I and *Not*I. The ends of the vector were then blunted and the vector religated; as a result pZLD is missing the restriction sites from *Sal*I to *Not*I. The 35S-polylinker-*nos* cassette was cloned into pSLJ75515 which had previously had superfluous restriction enzyme sites removed.

pAOVII and pSOVII (Fig. 2) were constructed using pAOV as backbone. The pZLD multiple cloning site was amplified by PCR using T7 and M13 primers and ligated into pAOV that had been previously modified to remove several conflicting restriction sites.

pKMB and pSMB (Fig. 2) were also constructed using pAOV as backbone. The pBluescript multiple cloning site was amplified by PCR using T3 and T7 primers and ligated into pAOV that had been previously modified to remove several conflicting restriction sites.

A detailed construction strategy flowchart is available and will be included with any vectors supplied.

Results and Discussion

pAOV allows the direct insertion of EST clones from the PRL2 library into a binary plasmid under the control of the 35S CaMV promoter and the *nos* terminator sequences. EST clones obtained from either the Arabidopsis Biological Research Center (ABRC, Ohio State) or the Nottingham Arabidopsis Stock Centre (NASC, Nottingham) as bacterial stab cultures can be grown, miniprepped and the inserts excised with *Sma*I at the 5' end and either *Xba*I or *Bam*HI at the 3' end. DNA bands purified from gels can be ligated into pAOV using a sticky/blunt ligation strategy (*Xba*I or *Bam*HI at 5' and *Ecl*136II at 3'). We have successfully prepared 80 antisense constructs with pAOV of which 40 have already been used to transform *Arabidopsis* by vacuum infiltration. We have just recently finished the construction of pAOV2 that contains several extra restriction sites in the polylinker allowing the use of either *Eco*RI or *Pst*I at the 5' side of the EST clone and *Xba*I, *Bam*HI or *Hin*dIII at the 3' end. pAOV2 will allow a more efficient sticky end ligation and provides several different possibilities for cloning.

pSOV was designed to allow the cloning of EST sequences from the *Arabidopsis* PRL2 library in sense orientation. The purpose of this vector is the production of plants

with extra copies of a particular gene to achieve either over-expression of the encoded protein or the overall down regulation of the gene by co-supression events. EST sequences can be cloned using the *Eco*RI site at the 5'end of the EST clone and the *Xba*I, *Bam*HI or *Hin*dII sites at the 3' end. We have also finished the construction of pSOV2 that contains an extra *Pst*I restriction site at the 5' end of the polylinker providing extra choices for the cloning strategy. There are no spurious ATG codons in either pSOV or pSOV2 that could interfere with the start of translation of the encoded protein.

Finally we have constructed pKMB and pSMB, two binary vectors also based on the pSLJ75515 backbone. Both vectors contain the 35S CaMV promoter and *nos* terminator sequences separated by the complete pBluescript polylinker in KS (pKMB) or SK (pSMB) orientations, including the T3 and T7 promoters. Please note that the *Xba*I site in pSMB occurs twice within the polylinker.

In conclusion, our collection of binary vectors will allow the quick production of sense and antisense constructs containing *Arabidopsis* ESTs in no more than a week after obtaining the *E. coli* cultures from any of the repository sites. They provide alternative strategies for the cloning, allowing for the possibility of the EST clone containing some of the designated restriction sites within its sequence. Using the vectors described here we routinely complete the process of obtaining an EST clone, inserting it into the binary vector in sense and/or antisense orientations, perform *in planta* transformation and select BASTA resistant seedlings in six weeks. Nevertheless, the use of these vectors is of course not limited to ESTs, especially for the pKMB and pSMB vectors.

All the vectors described in this paper are available upon request from the authors.

Acknowledgements

The authors would like to thank Dr. J.D.G. Jones and Dr. M.A. Botella for generously providing pSLJ75515.

References

Bechtold, N., J. Ellis and G. Pelletier. 1993. **In planta Agrobacterium mediated gene transfer by infiltration of adult Arabidopsis thaliana plants.** Comptes Rendus de l'Academie des Sciences Serie III Sciences de la Vie 316: 1194-1199.

Bevan, M. 1984. **Binary Agrobacterium vectors for plant transformation.** Nucleic Acids Research 12: 8711-8721.

De Block, M., J. Botterman, M. Vandewiele, J. Dockx, C. Thoen, V. Gossele, N.R. Movva, C. Thompson, M. Van Montagu and J. Leemans. 1987. **Engineering herbicide resistance in plants by expression of a detoxifying enzyme.** EMBO Journal 6: 2513-2518.

Jones, J.D., L. Shlumukov, F. Carland, J. English, S.R. Scofield, G.J. Bishop and K. Harrison. 1992. **Effective vectors for transformation, expression of heterologous genes, and assaying transposon excision in transgenic plants.** Transgenic Research 1: 285-297.

Lee, K.Y., J. Townsend, J. Tepperman, M. Black, C.F. Chui, B. Mazur, P. Dunsmuir and J. Bedbrook. 1988. **The molecular basis of sulfonylurea herbicide resistance in tobacco.** EMBO Journal 7: 1241-1248.

Newman, T., F.J. De-Bruijn, P. Green, K. Keegstra, H. Kende, L. McIntosh, J. Ohlrogge, N. Raikhel, S. Somerville, M. Thomashow, E. Retzel and C. Somerville. 1994. **Genes Galore: A Summary of Methods for Accessing Results from Large-Scale Partial Sequencing of Anonymous Arabidopsis cDNA Clones.** Plant Physiology 106: 1241-1255.

Sambrook, J., E.F. Fritsch and T. Maniatis. 1989. **Molecular Cloning: a Laboratory Manual**, 2nd ed. Cold Spring Harbor Laboratory Press, New York.

Thompson, C.J., N.R. Movva, R. Tizard, R. Crameri, J.E. Davies, M. Lauwereys and J. Botterman. 1987. **Characterization of the herbicide-resistance gene bar from Streptomyces hygroscopicus.** EMBO Journal 6: 2519-2524.

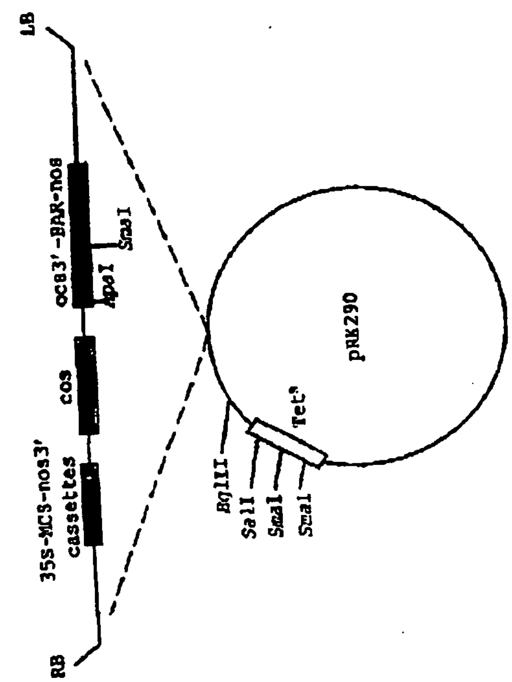
Figure Legends

Fig. 1. General map of the vectors described in this paper. The vector pSLJ75515 was used to introduce several cassettes (see Fig. 2) each containing the 35S CaMV promoter, a specific multiple cloning site (MCS) and *nos* terminator. pSLJ75515 is a derivative of pRK290 (Jones et al., 1992). The total size of all the binary vectors described in this paper is 28 kb.

Fig. 2. Construction of the different sense and antisense cassettes. The figure shows the most prominent details of the PRL2 library vector and the different cassettes constructed to insert EST clones in either sense or antisense orientation under the control of the CaMV 35S promoter and *nos* terminator. Restriction enzyme sites that can be used for the cloning of EST inserts are shown in bold. LB and RB represent T-DNA left and right border sequences respectively. The nopaline synthase promoter and terminators are represented as nos and nos3' respectively. 35S and ocs 3' indicates CaMV 35S promoter and octopine synthase 3' end respectively. BAR indicates the bialophos resistance gene, which encodes the enzyme phosphinotricin acetyltransferase (Thompson et al., 1987).

[1] restriction site not tested. * not unique.

FIG. 1


LB
OCS3'-BAR-nos
SmaI
ApoI
cos
35S-MCS-nos3'
cassettes
RB
pRK290
Tet
BglII
SalI
SmaI
SmaI

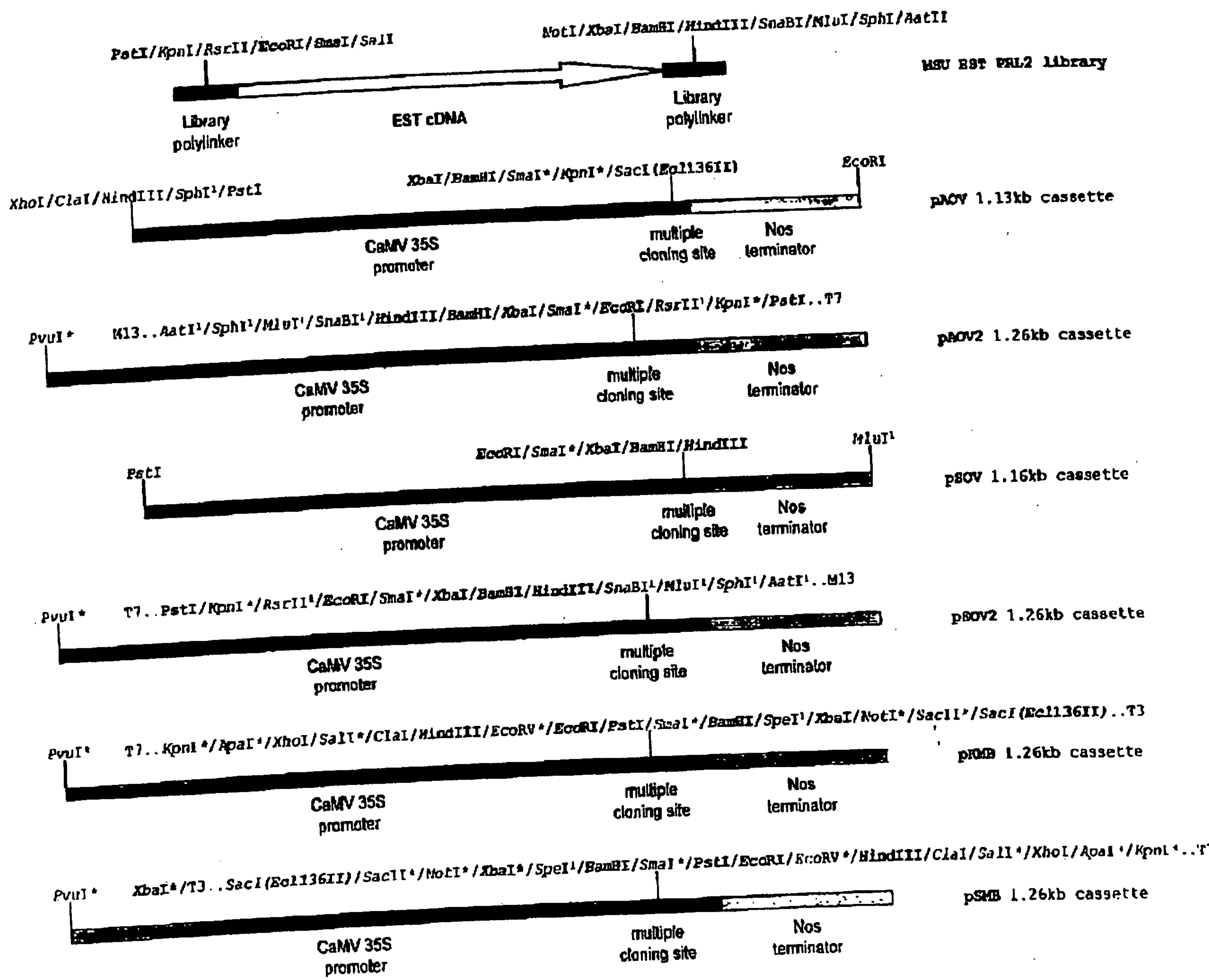

PstI/KpnI/RsrII/EcoRI/SmaI/SalI
NotI/XbaI/BamHI/HindIII/SnaBI/MluI/SphI/AatII
MSU EST PRL2 library
Library polylinker
EST cDNA
Library polylinker
XhoI/ClaI/HindIII/SphI¹/PstI
XbaI/BamHI/SmaI*/KpnI*/SacI(Ecl136II)
EcoRI
pAOV 1.13kb cassette
CaMV 35S promoter
multiple cloning site
Nos terminator
PvuI*
M13..AatI¹/SphI¹/MluI¹/SnaBI¹/HindIII/BamHI/XbaI/SmaI*/EcoRI/RsrII¹/KpnI*/PstI..T7
pAOV2 1.26kb cassette
CaMV 35S promoter
multiple cloning site
Nos terminator
PstI
EcoRI/SmaI*/XbaI/BamHI/HindIII
MluI¹
pSOV 1.16kb cassette
CaMV 35S promoter
multiple cloning site
Nos terminator
PvuI*
T7..PstI/KpnI*/RsrII¹/EcoRI/SmaI*/XbaI/BamHI/HindIII/SnaBI¹/MluI¹/SphI¹/AatI¹..M13
pSOV2 1.26kb cassette
CaMV 35S promoter
multiple cloning site
Nos terminator
PvuI*
T7..KpnI*/ApaI*/XhoI/SalI*/ClaI/HindIII/EcoRV*/EcoRI/PstI/SmaI*/BamHI/SpeI¹/XbaI/NotI*/SacII*/SacI(Ecl136II)..T3
pRMB 1.26kb cassette
CaMV 35S promoter
multiple cloning site
Nos terminator
PvuI*
XbaI*/T3..SacI(Ecl136II)/SacII*/NotI*/XbaI*/SpeI¹/BamHI/SmaI*/PstI/EcoRI/EcoRV*/HindIII/ClaI/SalI*/XhoI/ApaI*/KpnI*..T7
pSMB 1.26kb cassette
CaMV 35S promoter
multiple cloning site
Nos terminator

FIG. 2

WHAT IS CLAIMED IS:

1. A method of constructing a synthetic polynucleotide from which a protein is selectively expressible in a target cell of a plant, relative to another cell of the plant, said method comprising:

– selecting a first codon of a parent polynucleotide for replacement with a synonymous codon which has a higher translational efficiency in said target cell than in said other cell; and

– replacing said first codon with said synonymous codon to form said synthetic polynucleotide.

2. The method of claim 1, wherein said first codon and said synonymous codon are selected by:

– comparing translational efficiencies of individual codons in said target cell relative to said other cell; and

– selecting said first codon and said synonymous codon based on said comparison.

3. The method of claim 2, wherein the translational efficiency of an individual codon is measured by:

– introducing into said target cell and into said other cell, a synthetic construct comprising a reporter polynucleotide fused in frame with a tandem repeat of said individual codon, wherein said reporter polynucleotide encodes a reporter protein, and wherein said synthetic construct is operably linked to a regulatory polynucleotide; and

– comparing expression of said reporter protein in said target cell and in said other cell to determine the

translational efficiency of said individual codon in said target cell relative to said other cell.

4. The method of claim 3, further comprising:

– introducing the synthetic construct into a progenitor cell of a cell selected from the group consisting of said target cell and said other cell; and

– producing said target cell from said progenitor cell, wherein said cell contains said synthetic construct.

5. The method of claim 3, further comprising:

– introducing the synthetic construct into a progenitor cell of a cell selected from the group consisting of said target cell and said another cell; and

– growing a plant or part thereof from said progenitor cell, wherein said plant comprises said cell containing said synthetic construct.

6. The method of claim 3, further comprising: introducing the synthetic construct into an plant or part thereof such that said synthetic construct is introduced into said target cell or said other cell.

7. The method of claim 2, wherein said synonymous codon corresponds to a reporter construct from which the reporter protein is expressed in said target cell at a level that is at least 110% the level of the reporter protein that is expressed from the same reporter construct in said other cell.

8. The method of claim 2, wherein the translational efficiency of an individual codon is compared by measuring the abundance of an iso-tRNA corresponding to said individual codon in said target cell relative to said other cell.

9. The method of claim 8, wherein said synonymous codon corresponds to an iso-tRNA which is in higher abundance in said target cell relative to said other cell.

10. The method of claim 8, wherein selecting said first codon and said synonymous codon comprises:

– measuring abundance of different iso-tRNAs in said target cell relative to said other cell; and

– selecting said first codon and said synonymous codon based on said measurement, wherein said synonymous codon corresponds to an iso-tRNA which is in higher abundance in said target cell than in said other cell.

11. The method of claim 8, wherein said synonymous codon corresponds to an iso-tRNA that is present in said target cell at a level which is at least 110% of the level of the iso-tRNA that is present in said other cell.

12. The method of claim 1, wherein said synonymous codon is selected from the group consisting of (1) a codon used at relatively high frequency by genes of said target cell, (2) a codon used at relatively high frequency by genes of the plant, (3) a codon used at relatively low frequency by genes of said other cell, and (4) a codon used at relatively low frequency by genes of another organism.

13. The method of claim 1, wherein said first codon is selected from the group consisting of (a) a codon used at relatively high frequency by genes of said other cell, (b) a codon used at relatively low frequency by genes of said target cell, (c) a codon used at relatively low frequency by genes of the plant, and (d) a codon used at relatively high frequency by genes of another organism.

14. The method of claim 1 wherein said first codon and said synonymous codon are selected such that said protein is

expressed from said synthetic polynucleotide in said target cell at a level which is at least 110% of the level at which said protein is expressed from said parent polynucleotide in said target cell.

15. The method of claim 1, wherein said other cell is a precursor cell of said target cell.

16. The method of claim 1, wherein said other cell is a cell derived from said target cell.

17. A synthetic polynucleotide constructed according to the method of claim 1.

18. A method of selectively expressing a protein in a target cell of a plant, said method comprising:

- replacing a first codon of a parent polynucleotide encoding said protein with a synonymous codon to produce a synthetic polynucleotide having altered translational kinetics compared to said parent polynucleotide, such that said protein is expressible in said target cell, but such that said protein is not substantially expressible in another cell of the plant; and

- introducing into a cell selected from the group consisting of said target cell and a precursor of said target cell, said synthetic polynucleotide operably linked to a regulatory polynucleotide,

whereby said protein is selectively expressed in said target cell.

19. The method of claim 18, wherein said synonymous codon has a higher translational efficiency in said target cell than in said other cell.

20. A method of expressing a protein in a target cell of a plant from a first polynucleotide, said method comprising:

– introducing into a cell selected from the group consisting of said target cell and a precursor of said target cell, a second polynucleotide encoding an iso-tRNA, wherein said second polynucleotide is operably linked to a regulatory polynucleotide, and wherein said iso-tRNA is normally in relatively low abundance in said target cell and corresponds to a codon of said first polynucleotide, whereby said protein is expressed in said target cell.

21. A vector comprising the synthetic polynucleotide of claim 17.

22. A cell comprising the synthetic polynucleotide of claim 17.

23. A cell comprising the vector of claim 18.

24. A cell produced by the method of claim 20.

25. The method of claim 1, wherein said protein is not substantially expressible in said other cell.

26. The method of claim 2, wherein said tandem repeat comprises at least three copies of said individual codon.

27. A transgenic plant or plant part comprising a cell containing the synthetic polynucleotides of claim 17.

SEQUENCE LISTING

<110> The University of Queensland (all designated States except US)
Frazer, Ian Hector; Zhou, Jian; and Botella, Jose (US only)

<120> Plant Codon Usage

<130> POLYNUCLEOTIDE AND METHOD FOR SELECTIVELY EXPRESSING A PROTEIN IN A TARGET CELL OR TISSUE OF A PLANT

<130> Codon optimization

<140> Not yet assigned
<141> Herewith

<150> PP8078
<151> 1999-01-08

<160> 36

<170> PatentIn Ver. 2.0

<210> 1
<211> 762
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: GFP control construct

<220>
<221> CDS
<222> (31)..(747)

<400> 1
```
tttaagcttg gatcccaagg agatataaca atg agt aaa gga gaa gaa ctt ttc   54
                                 Met Ser Lys Gly Glu Glu Leu Phe
                                   1               5

act gga gtt gtc cca att ctt gtt gaa tta gat ggt gat gtt aat ggg   102
Thr Gly Val Val Pro Ile Leu Val Glu Leu Asp Gly Asp Val Asn Gly
     10                  15                  20

cac aaa ttt tct gtc agt gga gag ggt gaa ggt gat gca aca tac gga   150
His Lys Phe Ser Val Ser Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly
 25                  30                  35                  40

aaa ctt acc ctt aaa ttt att tgc act act gga aaa cta cct gtt cca   198
Lys Leu Thr Leu Lys Phe Ile Cys Thr Thr Gly Lys Leu Pro Val Pro
                 45                  50                  55

tgg cca aca ctt gtc act act ttc tct tat ggt gtt caa tgc ttt tca   246
Trp Pro Thr Leu Val Thr Thr Phe Ser Tyr Gly Val Gln Cys Phe Ser
             60                  65                  70

aga tac cca gat cat atg aag cgg cac gac ttc ttc aag agc gcc atg   294
Arg Tyr Pro Asp His Met Lys Arg His Asp Phe Phe Lys Ser Ala Met
         75                  80                  85
```

cct gag gga tac gtg cag gag agg acc atc ttc ttc aag gac gac ggg 342
Pro Glu Gly Tyr Val Gln Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly
90 95 100

aac tac aag aca cgt gct gaa gtc aag ttt gag gga gac acc ctc gtc 390
Asn Tyr Lys Thr Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu Val
105 110 115 120

aac agg atc gag ctt aag gga atc gat ttc aag gag gac gga aac atc 438
Asn Arg Ile Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile
125 130 135

ctc ggc cac aag ttg gaa tac aac tac aac tcc cac aac gta tac atc 486
Leu Gly His Lys Leu Glu Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile
140 145 150

atg gcc gac aag caa aag aac ggc atc aaa gcc aac ttc aag acc cgc 534
Met Ala Asp Lys Gln Lys Asn Gly Ile Lys Ala Asn Phe Lys Thr Arg
155 160 165

cac aac atc gaa gac ggc ggc gtg caa ctc gct gat cat tat caa caa 582
His Asn Ile Glu Asp Gly Gly Val Gln Leu Ala Asp His Tyr Gln Gln
170 175 180

aat act cca att ggc gat ggc cct gtc ctt tta cca gac aac cat tac 630
Asn Thr Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn His Tyr
185 190 195 200

ctg tcc aca caa tct gcc ctt tcg aaa gat ccc aac gaa aag aga gac 678
Leu Ser Thr Gln Ser Ala Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp
205 210 215

cac atg gtc ctt ctt gag ttt gta aca gct gct ggg att aca cat ggc 726
His Met Val Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Thr His Gly
220 225 230

atg gat gaa cta tac aaa taa gaattcctgc agaaa 762
Met Asp Glu Leu Tyr Lys
235

<210> 2
<211> 238
<212> PRT
<213> Artificial Sequence

<400> 2
Met Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val
1 5 10 15

Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu
20 25 30

Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys
35 40 45

Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Phe
50 55 60

```
Ser Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys Arg
65                  70                  75                  80

His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu Arg
                85                  90                  95

Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val
            100                 105                 110

Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Ile
        115                 120                 125

Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn
    130                 135                 140

Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn Gly
145                 150                 155                 160

Ile Lys Ala Asn Phe Lys Thr Arg His Asn Ile Glu Asp Gly Gly Val
                165                 170                 175

Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly Pro
            180                 185                 190

Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu Ser
        195                 200                 205

Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe Val
    210                 215                 220

Thr Ala Ala Gly Ile Thr His Gly Met Asp Glu Leu Tyr Lys
225                 230                 235


<210> 3
<211> 780
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: His(CAT)6 GFP construct

<220>
<221> CDS
<222> (31)..(765)

<400> 3
tttaagcttg gatcccaagg agatataaca atg cat cat cat cat cat cat agt      54
                                  Met His His His His His His Ser
                                  1               5

aaa gga gaa gaa ctt ttc act gga gtt gtc cca att ctt gtt gaa tta     102
Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val Glu Leu
    10                  15                  20

gat ggt gat gtt aat ggg cac aaa ttt tct gtc agt gga gag ggt gaa     150
Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu Gly Glu
25                  30                  35                  40

ggt gat gca aca tac gga aaa ctt acc ctt aaa ttt att tgc act act     198
```

```
Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys Thr Thr
                45                  50                  55

gga aaa cta cct gtt cca tgg cca aca ctt gtc act act ttc tct tat     246
Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Phe Ser Tyr
            60                  65                  70

ggt gtt caa tgc ttt tca aga tac cca gat cat atg aag cgg cac gac     294
Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys Arg His Asp
        75                  80                  85

ttc ttc aag agc gcc atg cct gag gga tac gtg cag gag agg acc atc     342
Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu Arg Thr Ile
    90                  95                  100

ttc ttc aag gac gac ggg aac tac aag aca cgt gct gaa gtc aag ttt     390
Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val Lys Phe
105                 110                 115                 120

gag gga gac acc ctc gtc aac agg atc gag ctt aag gga atc gat ttc     438
Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Ile Asp Phe
                125                 130                 135

aag gag gac gga aac atc ctc ggc cac aag ttg gaa tac aac tac aac     486
Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn Tyr Asn
            140                 145                 150

tcc cac aac gta tac atc atg gcc gac aag caa aag aac ggc atc aaa     534
Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn Gly Ile Lys
        155                 160                 165

gcc aac ttc aag acc cgc cac aac atc gaa gac ggc ggc gtg caa ctc     582
Ala Asn Phe Lys Thr Arg His Asn Ile Glu Asp Gly Gly Val Gln Leu
    170                 175                 180

gct gat cat tat caa caa aat act cca att ggc gat ggc cct gtc ctt     630
Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly Pro Val Leu
185                 190                 195                 200

tta cca gac aac cat tac ctg tcc aca caa tct gcc ctt tcg aaa gat     678
Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu Ser Lys Asp
                205                 210                 215

ccc aac gaa aag aga gac cac atg gtc ctt ctt gag ttt gta aca gct     726
Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe Val Thr Ala
            220                 225                 230

gct ggg att aca cat ggc atg gat gaa cta tac aaa taa gaattcctgc      775
Ala Gly Ile Thr His Gly Met Asp Glu Leu Tyr Lys
        235                 240                 245

agaaa                                                               780
```

<210> 4
<211> 244
<212> PRT
<213> Artificial Sequence

<400> 4

Met His His His His His His Ser Lys Gly Glu Glu Leu Phe Thr Gly
1 5 10 15

Val Val Pro Ile Leu Val Glu Leu Asp Gly Asp Val Asn Gly His Lys
20 25 30

Phe Ser Val Ser Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu
35 40 45

Thr Leu Lys Phe Ile Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro
50 55 60

Thr Leu Val Thr Thr Phe Ser Tyr Gly Val Gln Cys Phe Ser Arg Tyr
65 70 75 80

Pro Asp His Met Lys Arg His Asp Phe Phe Lys Ser Ala Met Pro Glu
85 90 95

Gly Tyr Val Gln Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr
100 105 110

Lys Thr Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg
115 120 125

Ile Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly
130 135 140

His Lys Leu Glu Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala
145 150 155 160

Asp Lys Gln Lys Asn Gly Ile Lys Ala Asn Phe Lys Thr Arg His Asn
165 170 175

Ile Glu Asp Gly Gly Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr
180 185 190

Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser
195 200 205

Thr Gln Ser Ala Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met
210 215 220

Val Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Thr His Gly Met Asp
225 230 235 240

Glu Leu Tyr Lys

<210> 5
<211> 780
<212> DNA
<213> Artificial Sequence

<220>
<221> CDS
<222> (31)..(765)

<220>
<223> Description of Artificial Sequence: His(CAC)$_6$ GFP construct

```
<400> 5
tttaagcttg gatcccaagg agatataaca atg cac cac cac cac cac cac agt   54
                                 Met His His His His His His Ser
                                  1               5

aaa gga gaa gaa ctt ttc act gga gtt gtc cca att ctt gtt gaa tta   102
Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val Glu Leu
    10                  15                  20

gat ggt gat gtt aat ggg cac aaa ttt tct gtc agt gga gag ggt gaa   150
Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu Gly Glu
25                  30                  35                  40

ggt gat gca aca tac gga aaa ctt acc ctt aaa ttt att tgc act act   198
Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys Thr Thr
                45                  50                  55

gga aaa cta cct gtt cca tgg cca aca ctt gtc act act ttc tct tat   246
Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Phe Ser Tyr
            60                  65                  70

ggt gtt caa tgc ttt tca aga tac cca gat cat atg aag cgg cac gac   294
Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys Arg His Asp
        75                  80                  85

ttc ttc aag agc gcc atg cct gag gga tac gtg cag gag agg acc atc   342
Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu Arg Thr Ile
    90                  95                  100

ttc ttc aag gac gac ggg aac tac aag aca cgt gct gaa gtc aag ttt   390
Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val Lys Phe
105                 110                 115                 120

gag gga gac acc ctc gtc aac agg atc gag ctt aag gga atc gat ttc   438
Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Ile Asp Phe
                125                 130                 135

aag gag gac gga aac atc ctc ggc cac aag ttg gaa tac aac tac aac   486
Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn Tyr Asn
            140                 145                 150

tcc cac aac gta tac atc atg gcc gac aag caa aag aac ggc atc aaa   534
Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn Gly Ile Lys
        155                 160                 165

gcc aac ttc aag acc cgc cac aac atc gaa gac ggc ggc gtg caa ctc   582
Ala Asn Phe Lys Thr Arg His Asn Ile Glu Asp Gly Gly Val Gln Leu
    170                 175                 180

gct gat cat tat caa caa aat act cca att ggc gat ggc cct gtc ctt   630
Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly Pro Val Leu
185                 190                 195                 200

tta cca gac aac cat tac ctg tcc aca caa tct gcc ctt tcg aaa gat   678
Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu Ser Lys Asp
                205                 210                 215

ccc aac gaa aag aga gac cac atg gtc ctt ctt gag ttt gta aca gct   726
Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe Val Thr Ala
            220                 225                 230
```

```
gct ggg att aca cat ggc atg gat gaa cta tac aaa taa gaattcctgc     775
Ala Gly Ile Thr His Gly Met Asp Glu Leu Tyr Lys
        235                 240                 245

agaaa                                                              780


<210> 6
<211> 244
<212> PRT
<213> Artificial Sequence

<400> 6
Met His His His His His His Ser Lys Gly Glu Glu Leu Phe Thr Gly
  1               5                  10                  15

Val Val Pro Ile Leu Val Glu Leu Asp Gly Asp Val Asn Gly His Lys
             20                  25                  30

Phe Ser Val Ser Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu
         35                  40                  45

Thr Leu Lys Phe Ile Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro
     50                  55                  60

Thr Leu Val Thr Thr Phe Ser Tyr Gly Val Gln Cys Phe Ser Arg Tyr
 65                  70                  75                  80

Pro Asp His Met Lys Arg His Asp Phe Phe Lys Ser Ala Met Pro Glu
                 85                  90                  95

Gly Tyr Val Gln Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr
            100                 105                 110

Lys Thr Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg
        115                 120                 125

Ile Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly
    130                 135                 140

His Lys Leu Glu Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala
145                 150                 155                 160

Asp Lys Gln Lys Asn Gly Ile Lys Ala Asn Phe Lys Thr Arg His Asn
                165                 170                 175

Ile Glu Asp Gly Gly Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr
            180                 185                 190

Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser
        195                 200                 205

Thr Gln Ser Ala Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met
    210                 215                 220

Val Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Thr His Gly Met Asp
225                 230                 235                 240

Glu Leu Tyr Lys
```

<210> 7
<211> 780
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Leu(CTA)$_6$ GFP construct

<220>
<221> CDS
<222> (31)..(765)

<400> 7
tttaagcttg gatcccaagg agatataaca atg cta cta cta cta cta cta agt 54
Met Leu Leu Leu Leu Leu Leu Ser
1 5

aaa gga gaa gaa ctt ttc act gga gtt gtc cca att ctt gtt gaa tta 102
Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val Glu Leu
10 15 20

gat ggt gat gtt aat ggg cac aaa ttt tct gtc agt gga gag ggt gaa 150
Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu Gly Glu
25 30 35 40

ggt gat gca aca tac gga aaa ctt acc ctt aaa ttt att tgc act act 198
Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys Thr Thr
45 50 55

gga aaa cta cct gtt cca tgg cca aca ctt gtc act act ttc tct tat 246
Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Phe Ser Tyr
60 65 70

ggt gtt caa tgc ttt tca aga tac cca gat cat atg aag cgg cac gac 294
Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys Arg His Asp
75 80 85

ttc ttc aag agc gcc atg cct gag gga tac gtg cag gag agg acc atc 342
Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu Arg Thr Ile
90 95 100

ttc ttc aag gac gac ggg aac tac aag aca cgt gct gaa gtc aag ttt 390
Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val Lys Phe
105 110 115 120

gag gga gac acc ctc gtc aac agg atc gag ctt aag gga atc gat ttc 438
Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Ile Asp Phe
125 130 135

aag gag gac gga aac atc ctc ggc cac aag ttg gaa tac aac tac aac 486
Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn Tyr Asn
140 145 150

tcc cac aac gta tac atc atg gcc gac aag caa aag aac ggc atc aaa 534
Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn Gly Ile Lys
155 160 165

gcc aac ttc aag acc cgc cac aac atc gaa gac ggc ggc gtg caa ctc 582

```
Ala Asn Phe Lys Thr Arg His Asn Ile Glu Asp Gly Gly Val Gln Leu
    170                 175                 180

gct gat cat tat caa caa aat act cca att ggc gat ggc cct gtc ctt      630
Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly Pro Val Leu
185                 190                 195                 200

tta cca gac aac cat tac ctg tcc aca caa tct gcc ctt tcg aaa gat      678
Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu Ser Lys Asp
                205                 210                 215

ccc aac gaa aag aga gac cac atg gtc ctt ctt gag ttt gta aca gct      726
Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe Val Thr Ala
            220                 225                 230

gct ggg att aca cat ggc atg gat gaa cta tac aaa taa gaattcctgc       775
Ala Gly Ile Thr His Gly Met Asp Glu Leu Tyr Lys
        235                 240                 245

agaaa                                                                780
```

```
<210> 8
<211> 244
<212> PRT
<213> Artificial Sequence

<400> 8
Met Leu Leu Leu Leu Leu Leu Ser Lys Gly Glu Glu Leu Phe Thr Gly
  1               5                  10                  15

Val Val Pro Ile Leu Val Glu Leu Asp Gly Asp Val Asn Gly His Lys
            20                  25                  30

Phe Ser Val Ser Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu
        35                  40                  45

Thr Leu Lys Phe Ile Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro
    50                  55                  60

Thr Leu Val Thr Thr Phe Ser Tyr Gly Val Gln Cys Phe Ser Arg Tyr
 65                  70                  75                  80

Pro Asp His Met Lys Arg His Asp Phe Phe Lys Ser Ala Met Pro Glu
                85                  90                  95

Gly Tyr Val Gln Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr
            100                 105                 110

Lys Thr Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg
        115                 120                 125

Ile Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly
    130                 135                 140

His Lys Leu Glu Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala
145                 150                 155                 160

Asp Lys Gln Lys Asn Gly Ile Lys Ala Asn Phe Lys Thr Arg His Asn
                165                 170                 175
```

```
Ile Glu Asp Gly Gly Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr
            180                 185                 190

Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser
        195                 200                 205

Thr Gln Ser Ala Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met
    210                 215                 220

Val Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Thr His Gly Met Asp
225                 230                 235                 240

Glu Leu Tyr Lys


<210> 9
<211> 780
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Leu(CTC)6 GFP construct

<220>
<221> CDS
<222> (31)..(765)

<400> 9
tttaagcttg gatcccaagg agatataaca atg ctc ctc ctc ctc ctc ctc agt     54
                                 Met Leu Leu Leu Leu Leu Leu Ser
                                   1               5

aaa gga gaa gaa ctt ttc act gga gtt gtc cca att ctt gtt gaa tta     102
Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val Glu Leu
    10                  15                  20

gat ggt gat gtt aat ggg cac aaa ttt tct gtc agt gga gag ggt gaa     150
Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu Gly Glu
 25                  30                  35                  40

ggt gat gca aca tac gga aaa ctt acc ctt aaa ttt att tgc act act     198
Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys Thr Thr
                 45                  50                  55

gga aaa cta cct gtt cca tgg cca aca ctt gtc act act ttc tct tat     246
Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Phe Ser Tyr
            60                  65                  70

ggt gtt caa tgc ttt tca aga tac cca gat cat atg aag cgg cac gac     294
Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys Arg His Asp
        75                  80                  85

ttc ttc aag agc gcc atg cct gag gga tac gtg cag gag agg acc atc     342
Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu Arg Thr Ile
    90                  95                 100

ttc ttc aag gac gac ggg aac tac aag aca cgt gct gaa gtc aag ttt     390
Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val Lys Phe
105                 110                 115                 120
```

```
gag gga gac acc ctc gtc aac agg atc gag ctt aag gga atc gat ttc     438
Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Ile Asp Phe
                125                 130                 135

aag gag gac gga aac atc ctc ggc cac aag ttg gaa tac aac tac aac     486
Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn Tyr Asn
            140                 145                 150

tcc cac aac gta tac atc atg gcc gac aag caa aag aac ggc atc aaa     534
Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn Gly Ile Lys
        155                 160                 165

gcc aac ttc aag acc cgc cac aac atc gaa gac ggc ggc gtg caa ctc     582
Ala Asn Phe Lys Thr Arg His Asn Ile Glu Asp Gly Gly Val Gln Leu
    170                 175                 180

gct gat cat tat caa caa aat act cca att ggc gat ggc cct gtc ctt     630
Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly Pro Val Leu
185                 190                 195                 200

tta cca gac aac cat tac ctg tcc aca caa tct gcc ctt tcg aaa gat     678
Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu Ser Lys Asp
                205                 210                 215

ccc aac gaa aag aga gac cac atg gtc ctt ctt gag ttt gta aca gct     726
Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe Val Thr Ala
            220                 225                 230

gct ggg att aca cat ggc atg gat gaa cta tac aaa taa gaattcctgc      775
Ala Gly Ile Thr His Gly Met Asp Glu Leu Tyr Lys
        235                 240                 245

agaaa                                                               780


<210> 10
<211> 244
<212> PRT
<213> Artificial Sequence

<400> 10
Met Leu Leu Leu Leu Leu Leu Ser Lys Gly Glu Glu Leu Phe Thr Gly
 1               5                  10                  15

Val Val Pro Ile Leu Val Glu Leu Asp Gly Asp Val Asn Gly His Lys
            20                  25                  30

Phe Ser Val Ser Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu
        35                  40                  45

Thr Leu Lys Phe Ile Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro
    50                  55                  60

Thr Leu Val Thr Thr Phe Ser Tyr Gly Val Gln Cys Phe Ser Arg Tyr
65                  70                  75                  80

Pro Asp His Met Lys Arg His Asp Phe Phe Lys Ser Ala Met Pro Glu
                85                  90                  95
```

Gly Tyr Val Gln Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr
100 105 110

Lys Thr Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg
115 120 125

Ile Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly
130 135 140

His Lys Leu Glu Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala
145 150 155 160

Asp Lys Gln Lys Asn Gly Ile Lys Ala Asn Phe Lys Thr Arg His Asn
165 170 175

Ile Glu Asp Gly Gly Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr
180 185 190

Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser
195 200 205

Thr Gln Ser Ala Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met
210 215 220

Val Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Thr His Gly Met Asp
225 230 235 240

Glu Leu Tyr Lys

<210> 11
<211> 780
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Leu(CTG)$_6$ GFP construct

<220>
<221> CDS
<222> (31)..(765)

<400> 11
tttaagcttg gatcccaagg agatataaca atg ctg ctg ctg ctg ctg ctg agt 54
Met Leu Leu Leu Leu Leu Leu Ser
1 5

aaa gga gaa gaa ctt ttc act gga gtt gtc cca att ctt gtt gaa tta 102
Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val Glu Leu
10 15 20

gat ggt gat gtt aat ggg cac aaa ttt tct gtc agt gga gag ggt gaa 150
Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu Gly Glu
25 30 35 40

ggt gat gca aca tac gga aaa ctt acc ctt aaa ttt att tgc act act 198
Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys Thr Thr
45 50 55

gga aaa cta cct gtt cca tgg cca aca ctt gtc act act ttc tct tat 246

```
Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Phe Ser Tyr
            60                  65                  70

ggt gtt caa tgc ttt tca aga tac cca gat cat atg aag cgg cac gac     294
Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys Arg His Asp
        75                  80                  85

ttc ttc aag agc gcc atg cct gag gga tac gtg cag gag agg acc atc     342
Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu Arg Thr Ile
    90                  95                  100

ttc ttc aag gac gac ggg aac tac aag aca cgt gct gaa gtc aag ttt     390
Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val Lys Phe
105                 110                 115                 120

gag gga gac acc ctc gtc aac agg atc gag ctt aag gga atc gat ttc     438
Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Ile Asp Phe
                125                 130                 135

aag gag gac gga aac atc ctc ggc cac aag ttg gaa tac aac tac aac     486
Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn Tyr Asn
            140                 145                 150

tcc cac aac gta tac atc atg gcc gac aag caa aag aac ggc atc aaa     534
Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn Gly Ile Lys
        155                 160                 165

gcc aac ttc aag acc cgc cac aac atc gaa gac ggc ggc gtg caa ctc     582
Ala Asn Phe Lys Thr Arg His Asn Ile Glu Asp Gly Gly Val Gln Leu
    170                 175                 180

gct gat cat tat caa caa aat act cca att ggc gat ggc cct gtc ctt     630
Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly Pro Val Leu
185                 190                 195                 200

tta cca gac aac cat tac ctg tcc aca caa tct gcc ctt tcg aaa gat     678
Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu Ser Lys Asp
                205                 210                 215

ccc aac gaa aag aga gac cac atg gtc ctt ctt gag ttt gta aca gct     726
Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe Val Thr Ala
            220                 225                 230

gct ggg att aca cat ggc atg gat gaa cta tac aaa taa gaattcctgc      775
Ala Gly Ile Thr His Gly Met Asp Glu Leu Tyr Lys
        235                 240                 245

agaaa                                                               780


<210> 12
<211> 244
<212> PRT
<213> Artificial Sequence

<400> 12
Met Leu Leu Leu Leu Leu Leu Ser Lys Gly Glu Glu Leu Phe Thr Gly
 1               5                  10                  15

Val Val Pro Ile Leu Val Glu Leu Asp Gly Asp Val Asn Gly His Lys
```

```
                20                  25                  30

Phe Ser Val Ser Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu
        35                  40                  45

Thr Leu Lys Phe Ile Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro
    50                  55                  60

Thr Leu Val Thr Thr Phe Ser Tyr Gly Val Gln Cys Phe Ser Arg Tyr
65                  70                  75                  80

Pro Asp His Met Lys Arg His Asp Phe Phe Lys Ser Ala Met Pro Glu
                85                  90                  95

Gly Tyr Val Gln Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr
            100                 105                 110

Lys Thr Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg
        115                 120                 125

Ile Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly
    130                 135                 140

His Lys Leu Glu Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala
145                 150                 155                 160

Asp Lys Gln Lys Asn Gly Ile Lys Ala Asn Phe Lys Thr Arg His Asn
                165                 170                 175

Ile Glu Asp Gly Gly Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr
            180                 185                 190

Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser
        195                 200                 205

Thr Gln Ser Ala Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met
    210                 215                 220

Val Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Thr His Gly Met Asp
225                 230                 235                 240

Glu Leu Tyr Lys


<210> 13
<211> 780
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Leu(CTT)6 GFP construct

<220>
<221> CDS
<222> (31)..(765)

<400> 13
tttaagcttg gatcccaagg agatataaca atg ctt ctt ctt ctt ctt ctt agt   54
                                 Met Leu Leu Leu Leu Leu Leu Ser
                                 1               5
```

```
aaa gga gaa gaa ctt ttc act gga gtt gtc cca att ctt gtt gaa tta     102
Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val Glu Leu
     10                  15                  20

gat ggt gat gtt aat ggg cac aaa ttt tct gtc agt gga gag ggt gaa     150
Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu Gly Glu
 25                  30                  35                  40

ggt gat gca aca tac gga aaa ctt acc ctt aaa ttt att tgc act act     198
Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys Thr Thr
                 45                  50                  55

gga aaa cta cct gtt cca tgg cca aca ctt gtc act act ttc tct tat     246
Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Phe Ser Tyr
             60                  65                  70

ggt gtt caa tgc ttt tca aga tac cca gat cat atg aag cgg cac gac     294
Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys Arg His Asp
         75                  80                  85

ttc ttc aag agc gcc atg cct gag gga tac gtg cag gag agg acc atc     342
Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu Arg Thr Ile
     90                  95                 100

ttc ttc aag gac gac ggg aac tac aag aca cgt gct gaa gtc aag ttt     390
Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val Lys Phe
105                 110                 115                 120

gag gga gac acc ctc gtc aac agg atc gag ctt aag gga atc gat ttc     438
Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Ile Asp Phe
                125                 130                 135

aag gag gac gga aac atc ctc ggc cac aag ttg gaa tac aac tac aac     486
Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn Tyr Asn
            140                 145                 150

tcc cac aac gta tac atc atg gcc gac aag caa aag aac ggc atc aaa     534
Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn Gly Ile Lys
        155                 160                 165

gcc aac ttc aag acc cgc cac aac atc gaa gac ggc ggc gtg caa ctc     582
Ala Asn Phe Lys Thr Arg His Asn Ile Glu Asp Gly Gly Val Gln Leu
    170                 175                 180

gct gat cat tat caa caa aat act cca att ggc gat ggc cct gtc ctt     630
Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly Pro Val Leu
185                 190                 195                 200

tta cca gac aac cat tac ctg tcc aca caa tct gcc ctt tcg aaa gat     678
Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu Ser Lys Asp
                205                 210                 215

ccc aac gaa aag aga gac cac atg gtc ctt ctt gag ttt gta aca gct     726
Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe Val Thr Ala
            220                 225                 230

gct ggg att aca cat ggc atg gat gaa cta tac aaa taa gaattcctgc      775
Ala Gly Ile Thr His Gly Met Asp Glu Leu Tyr Lys
        235                 240                 245
```

agaaa 780

<210> 14
<211> 244
<212> PRT
<213> Artificial Sequence

<400> 14
Met Leu Leu Leu Leu Leu Leu Ser Lys Gly Glu Glu Leu Phe Thr Gly
1 5 10 15

Val Val Pro Ile Leu Val Glu Leu Asp Gly Asp Val Asn Gly His Lys
20 25 30

Phe Ser Val Ser Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu
35 40 45

Thr Leu Lys Phe Ile Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro
50 55 60

Thr Leu Val Thr Thr Phe Ser Tyr Gly Val Gln Cys Phe Ser Arg Tyr
65 70 75 80

Pro Asp His Met Lys Arg His Asp Phe Phe Lys Ser Ala Met Pro Glu
85 90 95

Gly Tyr Val Gln Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr
100 105 110

Lys Thr Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg
115 120 125

Ile Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly
130 135 140

His Lys Leu Glu Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala
145 150 155 160

Asp Lys Gln Lys Asn Gly Ile Lys Ala Asn Phe Lys Thr Arg His Asn
165 170 175

Ile Glu Asp Gly Gly Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr
180 185 190

Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser
195 200 205

Thr Gln Ser Ala Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met
210 215 220

Val Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Thr His Gly Met Asp
225 230 235 240

Glu Leu Tyr Lys

<210> 15
<211> 780

<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Leu(TTA)$_6$ GFP construct

<220>
<221> CDS
<222> (31)..(765)

<400> 15

```
tttaagcttg gatcccaagg agatataaca atg tta tta tta tta tta tta agt   54
                                 Met Leu Leu Leu Leu Leu Leu Ser
                                  1               5

aaa gga gaa gaa ctt ttc act gga gtt gtc cca att ctt gtt gaa tta   102
Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val Glu Leu
     10                  15                  20

gat ggt gat gtt aat ggg cac aaa ttt tct gtc agt gga gag ggt gaa   150
Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu Gly Glu
 25                  30                  35                  40

ggt gat gca aca tac gga aaa ctt acc ctt aaa ttt att tgc act act   198
Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys Thr Thr
                 45                  50                  55

gga aaa cta cct gtt cca tgg cca aca ctt gtc act act ttc tct tat   246
Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Phe Ser Tyr
             60                  65                  70

ggt gtt caa tgc ttt tca aga tac cca gat cat atg aag cgg cac gac   294
Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys Arg His Asp
         75                  80                  85

ttc ttc aag agc gcc atg cct gag gga tac gtg cag gag agg acc atc   342
Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu Arg Thr Ile
     90                  95                 100

ttc ttc aag gac gac ggg aac tac aag aca cgt gct gaa gtc aag ttt   390
Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val Lys Phe
105                 110                 115                 120

gag gga gac acc ctc gtc aac agg atc gag ctt aag gga atc gat ttc   438
Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Ile Asp Phe
                125                 130                 135

aag gag gac gga aac atc ctc ggc cac aag ttg gaa tac aac tac aac   486
Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn Tyr Asn
            140                 145                 150

tcc cac aac gta tac atc atg gcc gac aag caa aag aac ggc atc aaa   534
Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn Gly Ile Lys
        155                 160                 165

gcc aac ttc aag acc cgc cac aac atc gaa gac ggc ggc gtg caa ctc   582
Ala Asn Phe Lys Thr Arg His Asn Ile Glu Asp Gly Gly Val Gln Leu
    170                 175                 180

gct gat cat tat caa caa aat act cca att ggc gat ggc cct gtc ctt   630
```

```
Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly Pro Val Leu
185                 190                 195                 200

tta cca gac aac cat tac ctg tcc aca caa tct gcc ctt tcg aaa gat     678
Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu Ser Lys Asp
                205                 210                 215

ccc aac gaa aag aga gac cac atg gtc ctt ctt gag ttt gta aca gct     726
Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe Val Thr Ala
            220                 225                 230

gct ggg att aca cat ggc atg gat gaa cta tac aaa taa gaattcctgc      775
Ala Gly Ile Thr His Gly Met Asp Glu Leu Tyr Lys
        235                 240                 245

agaaa                                                               780


<210> 16
<211> 244
<212> PRT
<213> Artificial Sequence

<400> 16
Met Leu Leu Leu Leu Leu Leu Ser Lys Gly Glu Glu Leu Phe Thr Gly
 1               5                  10                  15

Val Val Pro Ile Leu Val Glu Leu Asp Gly Asp Val Asn Gly His Lys
            20                  25                  30

Phe Ser Val Ser Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu
        35                  40                  45

Thr Leu Lys Phe Ile Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro
    50                  55                  60

Thr Leu Val Thr Thr Phe Ser Tyr Gly Val Gln Cys Phe Ser Arg Tyr
65                  70                  75                  80

Pro Asp His Met Lys Arg His Asp Phe Phe Lys Ser Ala Met Pro Glu
                85                  90                  95

Gly Tyr Val Gln Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr
            100                 105                 110

Lys Thr Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg
        115                 120                 125

Ile Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly
    130                 135                 140

His Lys Leu Glu Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala
145                 150                 155                 160

Asp Lys Gln Lys Asn Gly Ile Lys Ala Asn Phe Lys Thr Arg His Asn
                165                 170                 175

Ile Glu Asp Gly Gly Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr
            180                 185                 190
```

```
Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser
        195                 200                 205

Thr Gln Ser Ala Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met
    210                 215                 220

Val Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Thr His Gly Met Asp
225                 230                 235                 240

Glu Leu Tyr Lys


<210> 17
<211> 780
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Leu(TTG)6 GFP construct

<220>
<221> CDS
<222> (31)..(765)

<400> 17
tttaagcttg gatcccaagg agatataaca atg ttg ttg ttg ttg ttg ttg agt   54
                                 Met Leu Leu Leu Leu Leu Leu Ser
                                   1               5

aaa gga gaa gaa ctt ttc act gga gtt gtc cca att ctt gtt gaa tta   102
Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val Glu Leu
    10                  15                  20

gat ggt gat gtt aat ggg cac aaa ttt tct gtc agt gga gag ggt gaa   150
Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu Gly Glu
 25                  30                  35                  40

ggt gat gca aca tac gga aaa ctt acc ctt aaa ttt att tgc act act   198
Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys Thr Thr
                 45                  50                  55

gga aaa cta cct gtt cca tgg cca aca ctt gtc act act ttc tct tat   246
Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Phe Ser Tyr
             60                  65                  70

ggt gtt caa tgc ttt tca aga tac cca gat cat atg aag cgg cac gac   294
Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys Arg His Asp
         75                  80                  85

ttc ttc aag agc gcc atg cct gag gga tac gtg cag gag agg acc atc   342
Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu Arg Thr Ile
    90                  95                  100

ttc ttc aag gac gac ggg aac tac aag aca cgt gct gaa gtc aag ttt   390
Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val Lys Phe
105                 110                 115                 120

gag gga gac acc ctc gtc aac agg atc gag ctt aag gga atc gat ttc   438
Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Ile Asp Phe
                125                 130                 135
```

```
aag gag gac gga aac atc ctc ggc cac aag ttg gaa tac aac tac aac      486
Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn Tyr Asn
            140                 145                 150

tcc cac aac gta tac atc atg gcc gac aag caa aag aac ggc atc aaa      534
Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn Gly Ile Lys
        155                 160                 165

gcc aac ttc aag acc cgc cac aac atc gaa gac ggc ggc gtg caa ctc      582
Ala Asn Phe Lys Thr Arg His Asn Ile Glu Asp Gly Gly Val Gln Leu
    170                 175                 180

gct gat cat tat caa caa aat act cca att ggc gat ggc cct gtc ctt      630
Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly Pro Val Leu
185                 190                 195                 200

tta cca gac aac cat tac ctg tcc aca caa tct gcc ctt tcg aaa gat      678
Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu Ser Lys Asp
            205                 210                 215

ccc aac gaa aag aga gac cac atg gtc ctt ctt gag ttt gta aca gct      726
Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe Val Thr Ala
            220                 225                 230

gct ggg att aca cat ggc atg gat gaa cta tac aaa taa gaattcctgc       775
Ala Gly Ile Thr His Gly Met Asp Glu Leu Tyr Lys
        235                 240                 245

agaaa                                                                780
```

<210> 18
<211> 244
<212> PRT
<213> Artificial Sequence

<400> 18

```
Met Leu Leu Leu Leu Leu Leu Ser Lys Gly Glu Glu Leu Phe Thr Gly
  1               5                  10                  15

Val Val Pro Ile Leu Val Glu Leu Asp Gly Asp Val Asn Gly His Lys
            20                  25                  30

Phe Ser Val Ser Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu
        35                  40                  45

Thr Leu Lys Phe Ile Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro
    50                  55                  60

Thr Leu Val Thr Thr Phe Ser Tyr Gly Val Gln Cys Phe Ser Arg Tyr
65                  70                  75                  80

Pro Asp His Met Lys Arg His Asp Phe Phe Lys Ser Ala Met Pro Glu
                85                  90                  95

Gly Tyr Val Gln Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr
            100                 105                 110

Lys Thr Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg
```

```
        115                 120                 125

Ile Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly
    130                 135                 140

His Lys Leu Glu Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala
145                 150                 155                 160

Asp Lys Gln Lys Asn Gly Ile Lys Ala Asn Phe Lys Thr Arg His Asn
                165                 170                 175

Ile Glu Asp Gly Gly Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr
            180                 185                 190

Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser
        195                 200                 205

Thr Gln Ser Ala Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met
    210                 215                 220

Val Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Thr His Gly Met Asp
225                 230                 235                 240

Glu Leu Tyr Lys


<210> 19
<211> 780
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Lys(AAA)6 GFP construct

<220>
<221> CDS
<222> (31)..(765)

<400> 19
tttaagcttg gatcccaagg agatataaca atg aaa aaa aaa aaa aaa aaa agt   54
                                 Met Lys Lys Lys Lys Lys Lys Ser
                                   1               5

aaa gga gaa gaa ctt ttc act gga gtt gtc cca att ctt gtt gaa tta   102
Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val Glu Leu
     10                  15                  20

gat ggt gat gtt aat ggg cac aaa ttt tct gtc agt gga gag ggt gaa   150
Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu Gly Glu
 25                  30                  35                  40

ggt gat gca aca tac gga aaa ctt acc ctt aaa ttt att tgc act act   198
Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys Thr Thr
                 45                  50                  55

gga aaa cta cct gtt cca tgg cca aca ctt gtc act act ttc tct tat   246
Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Phe Ser Tyr
             60                  65                  70

ggt gtt caa tgc ttt tca aga tac cca gat cat atg aag cgg cac gac   294
```

Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys Arg His Asp
75 80 85

ttc ttc aag agc gcc atg cct gag gga tac gtg cag gag agg acc atc 342
Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu Arg Thr Ile
90 95 100

ttc ttc aag gac gac ggg aac tac aag aca cgt gct gaa gtc aag ttt 390
Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val Lys Phe
105 110 115 120

gag gga gac acc ctc gtc aac agg atc gag ctt aag gga atc gat ttc 438
Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Ile Asp Phe
125 130 135

aag gag gac gga aac atc ctc ggc cac aag ttg gaa tac aac tac aac 486
Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn Tyr Asn
140 145 150

tcc cac aac gta tac atc atg gcc gac aag caa aag aac ggc atc aaa 534
Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn Gly Ile Lys
155 160 165

gcc aac ttc aag acc cgc cac aac atc gaa gac ggc ggc gtg caa ctc 582
Ala Asn Phe Lys Thr Arg His Asn Ile Glu Asp Gly Gly Val Gln Leu
170 175 180

gct gat cat tat caa caa aat act cca att ggc gat ggc cct gtc ctt 630
Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly Pro Val Leu
185 190 195 200

tta cca gac aac cat tac ctg tcc aca caa tct gcc ctt tcg aaa gat 678
Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu Ser Lys Asp
205 210 215

ccc aac gaa aag aga gac cac atg gtc ctt ctt gag ttt gta aca gct 726
Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe Val Thr Ala
220 225 230

gct ggg att aca cat ggc atg gat gaa cta tac aaa taa gaattcctgc 775
Ala Gly Ile Thr His Gly Met Asp Glu Leu Tyr Lys
235 240 245

agaaa 780

<210> 20
<211> 244
<212> PRT
<213> Artificial Sequence

<400> 20
Met Lys Lys Lys Lys Lys Lys Ser Lys Gly Glu Glu Leu Phe Thr Gly
1 5 10 15

Val Val Pro Ile Leu Val Glu Leu Asp Gly Asp Val Asn Gly His Lys
20 25 30

Phe Ser Val Ser Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu
35 40 45

```
Thr Leu Lys Phe Ile Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro
     50                  55                  60

Thr Leu Val Thr Thr Phe Ser Tyr Gly Val Gln Cys Phe Ser Arg Tyr
 65                  70                  75                  80

Pro Asp His Met Lys Arg His Asp Phe Phe Lys Ser Ala Met Pro Glu
                 85                  90                  95

Gly Tyr Val Gln Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr
            100                 105                 110

Lys Thr Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg
        115                 120                 125

Ile Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly
    130                 135                 140

His Lys Leu Glu Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala
145                 150                 155                 160

Asp Lys Gln Lys Asn Gly Ile Lys Ala Asn Phe Lys Thr Arg His Asn
                165                 170                 175

Ile Glu Asp Gly Gly Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr
            180                 185                 190

Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser
        195                 200                 205

Thr Gln Ser Ala Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met
    210                 215                 220

Val Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Thr His Gly Met Asp
225                 230                 235                 240

Glu Leu Tyr Lys


<210> 21
<211> 780
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Lys(AAG)6 GFP construct

<220>
<221> CDS
<222> (31)..(765)

<400> 21
tttaagcttg gatcccaagg agatataaca atg aag aag aag aag aag aag agt    54
                                  Met Lys Lys Lys Lys Lys Lys Ser
                                    1               5

aaa gga gaa gaa ctt ttc act gga gtt gtc cca att ctt gtt gaa tta    102
Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val Glu Leu
     10                  15                  20
```

```
gat ggt gat gtt aat ggg cac aaa ttt tct gtc agt gga gag ggt gaa     150
Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu Gly Glu
 25                  30                  35                  40

ggt gat gca aca tac gga aaa ctt acc ctt aaa ttt att tgc act act     198
Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys Thr Thr
                 45                  50                  55

gga aaa cta cct gtt cca tgg cca aca ctt gtc act act ttc tct tat     246
Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Phe Ser Tyr
             60                  65                  70

ggt gtt caa tgc ttt tca aga tac cca gat cat atg aag cgg cac gac     294
Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys Arg His Asp
         75                  80                  85

ttc ttc aag agc gcc atg cct gag gga tac gtg cag gag agg acc atc     342
Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu Arg Thr Ile
     90                  95                 100

ttc ttc aag gac gac ggg aac tac aag aca cgt gct gaa gtc aag ttt     390
Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val Lys Phe
105                 110                 115                 120

gag gga gac acc ctc gtc aac agg atc gag ctt aag gga atc gat ttc     438
Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Ile Asp Phe
                125                 130                 135

aag gag gac gga aac atc ctc ggc cac aag ttg gaa tac aac tac aac     486
Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn Tyr Asn
            140                 145                 150

tcc cac aac gta tac atc atg gcc gac aag caa aag aac ggc atc aaa     534
Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn Gly Ile Lys
        155                 160                 165

gcc aac ttc aag acc cgc cac aac atc gaa gac ggc ggc gtg caa ctc     582
Ala Asn Phe Lys Thr Arg His Asn Ile Glu Asp Gly Gly Val Gln Leu
    170                 175                 180

gct gat cat tat caa caa aat act cca att ggc gat ggc cct gtc ctt     630
Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly Pro Val Leu
185                 190                 195                 200

tta cca gac aac cat tac ctg tcc aca caa tct gcc ctt tcg aaa gat     678
Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu Ser Lys Asp
                205                 210                 215

ccc aac gaa aag aga gac cac atg gtc ctt ctt gag ttt gta aca gct     726
Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe Val Thr Ala
            220                 225                 230

gct ggg att aca cat ggc atg gat gaa cta tac aaa taa gaattcctgc      775
Ala Gly Ile Thr His Gly Met Asp Glu Leu Tyr Lys
        235                 240                 245

agaaa                                                               780
```

<210> 22
<211> 244
<212> PRT
<213> Artificial Sequence

<400> 22
Met Lys Lys Lys Lys Lys Lys Ser Lys Gly Glu Glu Leu Phe Thr Gly
1   5   10   15

Val Val Pro Ile Leu Val Glu Leu Asp Gly Asp Val Asn Gly His Lys
            20   25   30

Phe Ser Val Ser Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu
        35   40   45

Thr Leu Lys Phe Ile Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro
    50   55   60

Thr Leu Val Thr Thr Phe Ser Tyr Gly Val Gln Cys Phe Ser Arg Tyr
65   70   75   80

Pro Asp His Met Lys Arg His Asp Phe Phe Lys Ser Ala Met Pro Glu
                85   90   95

Gly Tyr Val Gln Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr
            100   105   110

Lys Thr Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg
        115   120   125

Ile Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly
    130   135   140

His Lys Leu Glu Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala
145   150   155   160

Asp Lys Gln Lys Asn Gly Ile Lys Ala Asn Phe Lys Thr Arg His Asn
                165   170   175

Ile Glu Asp Gly Gly Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr
            180   185   190

Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser
        195   200   205

Thr Gln Ser Ala Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met
    210   215   220

Val Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Thr His Gly Met Asp
225   230   235   240

Glu Leu Tyr Lys

<210> 23
<211> 79
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: 3' oligonucleotide primer

<400> 23
tttctgcagg aattcttatt tgtatagttc atccatgcca tgtgtaatcc cagcagctgt 60
tacaaactca agaaggacc 79

<210> 24
<211> 107
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: 5' oligonucleotide primer incorporating His(CAC)$_6$

<400> 24
tttaagcttg gatcccaagg agatataaca atgcaccacc accaccacca cagtaaagga 60
gaagaacttt tcactggagt tgtcccaatt cttgttgaat tagatgg 107

<210> 25
<211> 107
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: 5' oligonucleotide primer incoporating His(CAT)$_6$

<400> 25
tttaagcttg gatcccaagg agatataaca atgcatcatc atcatcatca tagtaaagga 60
gaagaacttt tcactggagt tgtcccaatt cttgttgaat tagatgg 107

<210> 26
<211> 107
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: 5' oligonucleotide primer incorporating Leu(CTA)$_6$

<400> 26
tttaagcttg gatcccaagg agatataaca atgctactac tactactact aagtaaagga 60
gaagaacttt tcactggagt tgtcccaatt cttgttgaat tagatgg 107

<210> 27
<211> 107
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: 5' oligonucleotide primer incorporating Leu(CTC)$_6$

<400> 27
tttaagcttg gatcccaagg agatataaca atgctcctcc tcctcctcct cagtaaagga 60
gaagaacttt tcactggagt tgtcccaatt cttgttgaat tagatgg 107

<210> 28
<211> 107
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: 5' oligonucleotide primer incorporating Leu(CTG)$_6$

<400> 28
tttaagcttg gatcccaagg agatataaca atgctgctgc tgctgctgct gagtaaagga 60
gaagaacttt tcactggagt tgtcccaatt cttgttgaat tagatgg 107

<210> 29
<211> 107
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: 5' oligonucleotide primer incorporating Leu(CTT)$_6$

<400> 29
tttaagcttg gatcccaagg agatataaca atgcttcttc ttcttcttct tagtaaagga 60
gaagaacttt tcactggagt tgtcccaatt cttgttgaat tagatgg 107

<210> 30
<211> 107
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: 5' oligonucleotide primer incorporating Leu(TTA)$_6$

<400> 30
tttaagcttg gatcccaagg agatataaca atgttattat tattattatt aagtaaagga 60
gaagaacttt tcactggagt tgtcccaatt cttgttgaat tagatgg 107

<210> 31
<211> 107
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: 5' oligonucleotide primer incorporating Leu(TTG)$_6$

<400> 31
tttaagcttg gatcccaagg agatataaca atgttgttgt tgttgttgtt gagtaaagga 60
gaagaacttt tcactggagt tgtcccaatt cttgttgaat tagatgg 107

<210> 32
<211> 107
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: 5' oligonucleotide primer incorporating Lys(AAA)$_6$

<400> 32
tttaagcttg gatcccaagg agatataaca atgaaaaaaa aaaaaaaaaa aagtaaagga 60
gaagaacttt tcactggagt tgtcccaatt cttgttgaat tagatgg 107

<210> 33
<211> 107
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: 5' oligonucleotide primer incorporating Lys(AAG)$_6$

<400> 33
tttaagcttg gatcccaagg agatataaca atgaagaaga agaagaagaa gagtaaagga 60
gaagaacttt tcactggagt tgtcccaatt cttgttgaat tagatgg 107

<210> 34
<211> 107
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: 5' oligonucleotide primer control

<400> 34
tttaagcttg gatcccaagg agatataaca atgagtaaag gagaagaact tttcactgga 60
gttgtcccaa ttcttgttga attagatgg 89

<210> 35
<211> 730
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: m-gfp5-minus ER targeting (5') and retention signal (3')

<220>
<221> CDS
<222> (9)..(722)

<400> 35
```
gcctcgag agt aaa gga gaa gaa ctt ttc act gga gtt gtc cca att ctt   50
         Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu
           1               5                  10

gtt gaa tta gat ggt gat gtt aat ggg cac aaa ttt tct gtc agt gga    98
Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly
 15                  20                  25                  30

gag ggt gaa ggt gat gca aca tac gga aaa ctt acc ctt aaa ttt att   146
Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile
                 35                  40                  45

tgc act act gga aaa cta cct gtt cca tgg cca aca ctt gtc act act   194
```

```
Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr
            50                  55                  60

ttc tct tat ggt gtt caa tgc ttt tca aga tac cca gat cat atg aag     242
Phe Ser Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys
        65                  70                  75

cgg cac gac ttc ttc aag agc gcc atg cct gag gga tac gtg cag gag     290
Arg His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu
    80                  85                  90

agg acc atc ttc ttc aag gac gac ggg aac tac aag aca cgt gct gaa     338
Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu
95                  100                 105                 110

gtc aag ttt gag gga gac acc ctc gtc aac agg atc gag ctt aag gga     386
Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly
                115                 120                 125

atc gat ttc aag gag gac gga aac atc ctc ggc cac aag ttg gaa tac     434
Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr
            130                 135                 140

aac tac aac tcc cac aac gta tac atc atg gcc gac aag caa aag aac     482
Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn
        145                 150                 155

ggc atc aaa gcc aac ttc aag acc cgc cac aac atc gaa gac ggc ggc     530
Gly Ile Lys Ala Asn Phe Lys Thr Arg His Asn Ile Glu Asp Gly Gly
    160                 165                 170

gtg caa ctc gct gat cat tat caa caa aat act cca att ggc gat ggc     578
Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly
175                 180                 185                 190

cct gtc ctt tta cca gac aac cat tac ctg tcc aca caa tct gcc ctt     626
Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu
                195                 200                 205

tcg aaa gat ccc aac gaa aag aga gac cac atg gtc ctt ctt gag ttt     674
Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe
            210                 215                 220

gta aca gct gct ggg att aca cat ggc atg gat gaa cta tac aaa tga     722
Val Thr Ala Ala Gly Ile Thr His Gly Met Asp Glu Leu Tyr Lys
        225                 230                 235

gagctccg                                                            730


<210> 36
<211> 237
<212> PRT
<213> Artificial Sequence

<400> 36
Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val Glu
 1               5                  10                  15

Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu Gly
```

```
                20                  25                  30

Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys Thr
            35                  40                  45

Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Phe Ser
        50                  55                  60

Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys Arg His
65                  70                  75                  80

Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu Arg Thr
                85                  90                  95

Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val Lys
            100                 105                 110

Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Ile Asp
        115                 120                 125

Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn Tyr
    130                 135                 140

Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn Gly Ile
145                 150                 155                 160

Lys Ala Asn Phe Lys Thr Arg His Asn Ile Glu Asp Gly Gly Val Gln
                165                 170                 175

Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly Pro Val
            180                 185                 190

Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu Ser Lys
        195                 200                 205

Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe Val Thr
    210                 215                 220

Ala Ala Gly Ile Thr His Gly Met Asp Glu Leu Tyr Lys
225                 230                 235
```

INTERNATIONAL SEARCH REPORT	International application No. PCT/AU 00/00007

A. CLASSIFICATION OF SUBJECT MATTER

Int Cl[7]: C12N 15/29

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
SEE ELECTRONIC DATABASE BOX BELOW

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
SEE ELECTRONIC DATABASE BOX BELOW

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Chem Abs, Medline, WPID/Search terms: gene expression, synthetic (polynucleotide or nucleic acid), selective expression or targetting, codon, synonymous, replace, transfer, bias, usage.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	81999/98 (University of Queensland) 21 January 1999	

☐ Further documents are listed in the continuation of Box C ☐ See patent family annex

* Special categories of cited documents:
- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier application or patent but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed
- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- "&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
10 February 2000	18 FEB 2000

Name and mailing address of the ISA/AU	Authorized officer
AUSTRALIAN PATENT OFFICE PO BOX 200, WODEN ACT 2606, AUSTRALIA E-mail address: pct@ipaustralia.gov.au Facsimile No. (02) 6285 3929	PHILIPPA WYRDEMAN Telephone No.: (02) 6283 2554